Do Attitudes Toward Thinking Influence the Benefit of Memory Strategies? Exploring the Relationships Between Need for Cognition, the Drawing Effect, the Generation Effect, and the Production Effect in Free Recall

by © Justine Yick

A Thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science in Experimental Psychology

Department of Psychology

Faculty of Science

Memorial University of Newfoundland

August 2024

St. John’s, Newfoundland and Labrador

Abstract

Three experiments explored whether Need for Cognition (NFC) affects the memory benefits associated with drawing, generating, and producing while studying. In three experiments, undergraduate students studied a list of words while implementing a designated encoding strategy, followed by a free recall test. Experiment 1 had participants study half of the words by drawing and the other half by writing. In Experiment 2, participants generated half of the words from a definition and typed the other half. Finally, in Experiment 3, participants read half of the words aloud while studying, and read the remaining words silently. The three experiments replicated the memory benefits associated with the drawing effect, generation effect, and production effect (i.e., better recall from the more elaborative strategy), but NFC did not significantly predict the magnitude of these memory benefits or overall recall performance. These findings suggest that these encoding strategies may aid memory regardless of NFC level.

Keywords: memory, generation effect, production effect, drawing effect, Need for Cognition

General Summary

People often use study strategies to help them remember. These strategies can range from creating study material, to reading aloud, to drawing pictures – all well-established tasks that have been shown to improve memory. The present experiments investigated whether an individual's tendency to enjoy thinking (called "Need for Cognition") was associated with how well these study strategies improved memory for each participant. Three study strategies for learning a list of words were explored in three separate experiments: 1) drawing while studying; 2) guessing study material from clues; 3) reading aloud while studying. After participants finished their respective memory tasks, they completed a questionnaire to measure their Need for Cognition. I found that Need for Cognition did not appear to have a definitive relationship with the effectiveness of any of the study strategies, suggesting that these tasks may help memory regardless of one's tendency to enjoy thinking.

Acknowledgements

I would first like to thank Dr. Kathleen Hourihan for your guidance, patience, and support throughout this project. I would also like to thank my other supervisory committee members, Dr. Jonathan Fawcett and Dr. Heath Matheson, for both their insights on the statistical analyses, as well as their keen advice during the review process. Thank you to the students of the Metacognition Lab, the Neurofog Lab, and the CREATE Lab, for always fostering a community of support. Outside of the lab, I would also like to thank my friends and family: Kate Bartlett, Yimiao Zhao, Noah Pevie, Allan Yick, and Jonathan Yick (as well as Biscuit, Louie, and Luna). Thank you for all of your warmth and care during this time. You all helped keep my head above water during long data collection days and even longer thesis writing nights. Finally, I would like to give special thanks to my parents, James and Joyce Yick, for always being there for me.

Table of Contents

List of Figures

List of Tables

List of Appendices

Chapter 1: Introduction

Whether one is attempting to dedicate a new name to memory or is preparing for a final exam, study strategies are frequently employed to help us remember. In the lab, research has shown that creating study material (i.e., generation; Slamecka & Graf, 1978), reading study items aloud (i.e., production; MacLeod et al., 2010), and sketching a representation of to-be-remembered words (i.e., drawing; Wammes et al., 2016) are all tasks that reliably benefit memory. Moreover, differences in memory performance appear to be associated with certain personality traits. For instance, the Big Five model is a recognized framework of five personality traits: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Costa & McCrae, 1992). Within this framework, Openness to Experience differentiates intellectually curious, creative, and imaginative people from more cautious, withdrawn individuals. Concerning memory, individuals who score high in Openness to Experience also tend to show greater recall performance than those with lower openness scores (Pinard et al., 2023; Soubelet & Salthouse, 2010; Taconnat et al., 2022). But although memory performance has been explored with the Big Five personality traits, there has yet to be research concerning the relationship between memory and another individual difference known as Need for Cognition (NFC), or how much one enjoys thinking. As such, the present thesis sought to explore whether some people benefit more from drawing, generating, and producing in comparison to others based on their level of Need for Cognition (NFC).

Need for Cognition (NFC) is an individual difference characterized by how much one enjoys and engages in thinking (Cacioppo & Petty, 1982). Individuals high in NFC enjoy thinking and cognitive challenges more than their low-NFC counterparts. But in addition to differing in cognitive attitudes, high-NFC and low-NFC individuals also differ in terms of intrinsic and extrinsic motivation (Amabile et al., 1994), as well as in problem-solving style (Eigenberger et al., 2006; Elphinstone et al., 2014) and ability (Berzonsky & Sullivan, 1992; Heppner et al., 1983). Of particular relevance for this thesis is the finding that high-NFC individuals tend to recall more than their low-NFC counterparts (Cacioppo et al., 1983; Leding, 2011; Soubelet & Salthouse, 2010; Wootan & Leding, 2015).

How does NFC relate to recall performance? One potential explanation is that the inclination and positive opinion toward the act of thinking that defines NFC also facilitates a relationship between this personality trait, elaborative encoding, and memory performance. Specifically, the attention toward an item's meaning, and how it relates to prior knowledge, is known as elaborative encoding. Elaborative encoding has been shown to increase memory performance when compared to simply focusing on less meaningful aspects of an item, such as the shape of a word's font or whether it contains a particular letter (Craik & Lockhart, 1972). High-NFC individuals appear to spontaneously engage in more elaborative processing during encoding compared to low-NFC counterparts (Leding, 2011; Wootan & Leding, 2015). Specifically, NFC has been examined in association with performance in the Deese-Roediger-McDermott (DRM) paradigm (Graham, 2006; Leding, 2011; Wootan & Leding, 2015). Namely, the DRM paradigm measures false memory susceptibility (Roediger & McDermott, 1995). Participants are shown a list of words that are closely related (e.g., *nurse, patient, hospital*) to a critical lure that is *not* presented (e.g., *doctor*). Then when they are tested on their memory, participants tend to falsely report that they studied the critical lure. As such, the DRM paradigm can measure both true recall (memory for the studied list items) and false recall (false memory of the critical lures). False recall is thought to increase when individuals are more likely to engage in semantic processing, as they are then more likely to unintentionally think of the critical lure while studying the given list (Gallo & Roediger, 2002). High-NFC participants have demonstrated higher rates of both true and false recall in the DRM paradigm (Leding, 2011; Wootan & Leding, 2015). As such, Leding and colleagues suggest that because high-NFC individuals enjoy effortful thinking *more than* their low-NFC counterparts, those high in NFC are also more likely to voluntarily consider an item's meaning when attempting to memorize it. This inclination toward deeper processing then leads to greater recall performance.

Although it has been shown that NFC impacts memory performance, it has yet to be seen whether or not this personality trait influences the relative effectiveness of various study strategies. The effectiveness of study strategies (e.g., drawing; Wammes et al., 2016) is tested by comparing memory performance following two encoding conditions. One condition asks participants to use an active strategy (e.g., to draw) while studying a given set of stimuli. Meanwhile, the other condition typically consists of a baseline encoding strategy, such as writing. This baseline condition provides a comparative measure of memory performance. In contrasting the mnemonic outcomes of the active and baseline encoding conditions, one generally sees that the active memory strategy outperforms the baseline condition. This difference in memory performance between the two encoding conditions is what is known as the memory benefit of the active encoding strategy (also known as the memory effect).

Consequently, this thesis aims to examine whether individuals who differ in NFC also experience differences in the magnitude of the memory benefits associated with various active memory strategies. As such, the present thesis will first provide an overview of the NFC construct, with an emphasis on its role in memory research. This will be followed by a review of elaborative encoding and the levels of processing theory, as well as focuses on the drawing effect

(Wammes et al., 2016), generation effect (Slamecka & Graf, 1978), and production effect (MacLeod et al., 2010) paradigms.

Need for Cognition

The Construct and Scale

Need for Cognition (NFC) refers to one's tendency to partake in, and enjoy, effortful thinking and deliberation (Cacioppo & Petty, 1982). This has been defined as a stable trait that reflects individual differences in engagement with cognitive activities. Specifically, the NFC construct classifies individuals on a continuum ranging from *misers*, who are low in NFC, and *cognizers*, who demonstrate high NFC. Individuals who are high in NFC are intrinsically motivated to think deeply, finding satisfaction in expending mental effort. On the other hand, individuals low in NFC are more likely to consider this kind of deliberation to be bothersome and unrewarding. As such, they may only engage in it when provided with external incentives. For instance, in their original study, Cacioppo and Petty (1982) found that university students who scored high in NFC were also more likely to rate a complex number-circling task as enjoyable when compared to individuals who scored low on the scale. This is despite the fact that both groups rated the task as mentally effortful.

The notion that people show individual differences in their attitudes toward thinking originates from dual-process and dual-systems theories of judgment (Petty et al., 2009). These theories posit that there are two ways that people think, particularly when making judgements and decisions, as well as when learning. One way is quick, unconscious, and intuitive while the other is slower, more deliberative, and intentional (see Evans & Stanovich, 2013 for review). Regarding NFC, it has been shown that, when absent of any external incentive, low-NFC individuals are more likely to rely upon the intuitive way of thinking. This can be seen in how low-NFC individuals are more likely to engage in the use of stereotypes and simple cues when conducting judgments (Carter et al., 2006; Haugtvedt et al., 1992). Meanwhile, high-NFC individuals are more likely to consider all relevant information when making decisions (Cacioppo et al., 1983), showing an inclination toward more deliberative thinking.

Cacioppo and Petty (1982) originally adapted the NFC construct from Cohen et al. (1955), who first defined NFC as "…a need to structure relevant situations in meaningful, integrated ways…" as well as "…a need to understand" (p. 291). As Cohen et al.'s (1955) definition did not involve a formal measure of NFC, Cacioppo and Petty (1982) sought to devise a scale for the construct. In their first experiment, the authors recruited university professors, to represent individuals with high levels of NFC, as well as assembly line workers, to represent individuals with low NFC. Cacioppo and Petty (1982) then gave all of the participants a 45-item questionnaire, with these items conceptualized to represent Cohen et al.'s (1955) definition of NFC. The participants rated their agreement with the items using a 9-point Likert scale (+4 = very strong agreement, -4 = very strong disagreement). Out of the original 45 items, the two groups' scores significantly differed on 34 of them. These 34 items represented a single dominant factor: how much people enjoyed and engaged in thinking. Cacioppo and Petty (1982) then used this factor structure to create the original 34-item NFC scale. Subsequently, in their follow-up experiments, the authors replicated their original factor structure in a more restricted sample of introductory psychology students – demonstrating that university students also appeared to differ in their levels of NFC.

Although the original NFC scale consists of 34 items, the most commonly used version is the revised, 18-item NFC scale (NFC-18; Cacioppo et al., 1984). This shorter scale is characterized by the same single factor solution as the 34-item scale and asks participants to rate their agreement with 18 different statements. These items represent a mixture of low-NFC attitudes (e.g., "*It's enough for me that something gets the job done; I don't care how or why it works.*") as well as high-NFC ones (e.g., "*I like to have the responsibility of handling a situation that requires a lot of thinking.*"). The scale items are most commonly rated using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), although prior research has also used the original 9-point Likert scale (+4 = very strong agreement, -4 = very strong disagreement) that was employed by the 34-item NFC scale.

Regarding psychometric properties, the NFC-18 has demonstrated high internal consistency, with past research finding Cronbach's alphas above .85 (Cacioppo et al., 1984; see Cacioppo et al., 1996 for review). More modern explorations have also supported the internal consistency of the NFC-18, with Hussey and Hughes (2020) reporting a Cronbach's alpha of .889, a McDonald's omega total of .889, and a McDonald's omega hierarchical of .885. In addition to its internal consistency, the scale has also shown good test-retest reliability, with Sadowski and Gulgoz (1993) finding a test-retest correlation of .88, as well as Hussey and Hughes (2020) reporting a correlation of .85. Moreover, the NFC-18 has exhibited good measurement invariance – showing no significant differences across gender and age groups (Cacioppo et al., 1984; Hussey & Hughes, 2020).

In What Other Ways Do Misers Differ from Cognizers?

The NFC construct can be further explained by its associations with other measures. For instance, as NFC represents one's attitude and inclination toward thinking, this individual difference is also related to variations in information processing and problem-solving preferences. For instance, the Epistemic Preference Indicator (EPI) measures two epistemic styles: "default position" and "intellective position" (Eigenberger et al., 2006; Elphinstone et al., 2014). The “default position” is defined as a preference for cognitive strategies that require minimal effort (e.g., heuristics), while the “intellective position” represents a preference for more deliberative processing when it comes to thinking and judgment. In line with the NFC construct, NFC scores have a significant positive relationship with scores for the “intellective position” of the EPI (Eigenberger et al., 2006; Elphinstone et al., 2014). Conversely, NFC scores have a significant negative correlation with scores for the “default position”. In other words, individuals high in NFC are more likely to prefer elaborative forms of thinking and judgment, while individuals low in NFC have a greater tendency to prefer more effortless cognitive strategies.

In a similar vein, high-NFC participants tend to self-report a greater degree of problem-solving abilities than low-NFC participants (Berzonsky & Sullivan, 1992; Heppner et al., 1983). For instance, the Problem-Solving Inventory (PSI) is a measure of both problem-solving ability and style, with high scores classifying individuals as “problem solvers” (Heppner & Petersen, 1982). Heppner et al. (1983) found that individuals who scored high in NFC were also more likely to score higher on the Problem-Solving Inventory, compared to individuals who scored low in NFC. Likewise, low NFC scores have been associated with the tendency to ignore problems and decision-making (Berzonsky & Sullivan, 1992).

NFC also has a positive relationship with intelligence. But rather than NFC directly measuring cognitive ability, it is instead thought that individuals with high cognitive capacity are often rewarded for using these skills. Cacioppo et al., (1996) suggest that such individuals may receive both social reinforcement and constructive feedback that bolsters feelings of competency in cognitive skills. These rewards then foster an intrinsic enjoyment toward thinking which defines NFC (Cacioppo et al., 1996). On the other hand, individuals who score lower on intelligence measures may struggle more with cognitive tasks, resulting in a decrease in engagement and enjoyment (Cacioppo et al., 1996). This can be seen when Cacioppo and Petty (1982) originally found that individuals with high NFC also had higher self-reported American College Testing Program Exam scores than their low-NFC counterparts, a finding that has also been replicated in later research (Cacioppo et al., 1984; Olson et al., 1984; Petty & Jarvis, 1996). Additionally, Hill et al. (2013) found that NFC was positively associated with the Wechsler Adult Intelligence Scale and the Raven's Advanced Progressive Matrices.

People who are high in NFC also differ from low-NFC individuals with regards to motivation in other areas. Namely, NFC correlates with other scales that measure intrinsic motivation (Amabile et al., 1994; Olsen et al., 1984). For instance, Amabile et al. (1994) administered the Work Preference Inventory (WPI), a measure of both intrinsic and extrinsic motivation, in addition to assessing participants' NFC. They found that for participants who scored low in NFC, they also tended to score low in intrinsic motivation and high in extrinsic motivation. The inverse was found for the high-NFC participants, with a greater tendency to be intrinsically motivated rather than fueled by external factors.

Finally, in line with NFC capturing a tendency to enjoy deliberation and reasoning, high NFC scores do not appear to be associated with strong, negative emotions. NFC is not significantly correlated with test anxiety (Cacioppo & Petty, 1982). It is also not significantly correlated with the tendency to experience intense emotions (Petty & Jarvis, 1996), nor with anger (Olson et al., 1984), nor is it significantly related to how people use emotions to guide their social behaviour (Booth-Butterfield & Booth-Butterfield, 1990). However, it does appear to have a significant negative relationship with both state and trait anxiety, with Olsen et al. (1984) reporting that as NFC scores increased, anxiety scores on the State-Trait Personality Inventory decreased. Similarly, a recent meta-analysis by Zerna et al. (2024) found that NFC also has a negative relationship with self-consciousness, with low-NFC individuals being more likely to demonstrate self-consciousness than their high-NFC counterparts.

In sum, NFC represents how much one enjoys thinking and engaging in complex cognitive tasks. As such, it is unsurprising that individuals high in NFC are also associated with higher cognitive abilities – this construct is positively correlated with more effortful cognitive processing styles (Eigenberger et al., 2006; Elphinstone et al., 2014), problem-solving abilities (Berzonsky & Sullivan, 1992; Heppner et al., 1983), and intelligence indicator scores (Cacioppo et al., 1984; Hill et al., 2013; Olson et al., 1984; Petty & Jarvis, 1996). These positive relationships could occur for multiple reasons. People who are adept at complex thinking may be more likely to develop an intrinsic motivation toward thinking due to the social reinforcement of cognitive achievement (Cacioppo et al., 1996). Conversely, individuals who face more difficulty with cognitive tasks might feel more negatively about complex thinking. Another factor involves inherent levels of motivation. It appears that individuals high in NFC are more intrinsically motivated in their work than their low-NFC counterparts (Amabile et al., 1994; Olsen et al., 1984). As such, it is also possible that high-NFC individuals are more likely to develop intrinsic motivation than their low-NFC counterparts – a trait that contributes to how motivated one might feel toward effortful cognition. Finally, individuals high in NFC appear to interact less with strong, negative emotions – with low-NFC individuals being more likely to feel anxiety and self-consciousness than their high-NFC counterparts (Olsen et al., 1984; Zerna et al., 2024).

Elaborative Encoding and Levels of Processing Theory

Individuals high in NFC have also demonstrated more elaborative encoding than low-NFC counterparts (Leding, 2011; Wootan & Leding, 2015). Elaborative encoding (also known as elaborative rehearsal and relational rehearsal) is generally defined as “thinking about what the to-be-remembered items mean and how they're related to one another and to other things you already know" (p. 215; Reisberg, 2018). In other words, elaborative encoding involves thinking about the meaning of a given item as one tries to remember it. Assessing the meaning of study material and thinking about how it relates to prior knowledge has been shown to benefit memory more than just focusing on the perceptual features of the item itself (Craik & Tulving, 1975; Hyde & Jenkins, 1969; Slamecka & Graf, 1978). Elaborative encoding contrasts maintenance rehearsal, where one simply focuses on the to-be-remembered stimulus without connecting it to any prior knowledge (Baddeley et al., 2020; Reisberg, 2018).

Elaborative encoding is an example of deep processing (also known as semantic processing). Deep processing contrasts with shallow processing, which refers to thinking about the specific perceptual features of an item (such as the way the letters in a word look or the way the word sounds) rather than the item's meaning or its relationship to other pieces of information. As such, the levels of processing theory suggests that deep processing results in stronger memory traces with more retrieval cues, increasing the likelihood of retention (Craik & Lockhart, 1972).

For instance, Craik and Tulving (1975) showed that different levels of processing in incidental learning impact memory. Specifically, participants were shown a series of words and, without being told they were in a memory experiment, were asked to process the material in various ways. They were asked to either indicate whether the words were in a capitalized or lowercase font, whether the words rhymed with a given cue word, or to assess the meaning of the words by confirming if they fit into a given sentence. In a surprise recall test, participants remembered the most words when they paid attention to the meaning of those words during the initial exposure phase. In other words, participants demonstrated better memory performance when they engaged in deeper processing at encoding, compared to more shallow processing. As such, this difference in memory performance between deeply processed items and shallowly processed items is known as the levels of processing effect.

In other words, elaborative encoding – or the act of connecting a to-be-remembered item with prior knowledge – has been shown to benefit memory (Craik & Tulving, 1975; Hyde & Jenkins, 1969; Slamecka & Graf, 1978). This elaborative encoding is thought to result in greater memory performance as these associations create more internal cues to facilitate a greater chance of later retrieval (Craik & Lockhart, 1972). As will be discussed in the following section, NFC has been associated with the tendency to engage in elaborative encoding (Leding, 2011; Wootan & Leding, 2015), leading to greater recall performance in high-NFC individuals (Cacioppo et al., 1983; Leding, 2011; Soubelet & Salthouse, 2010; Wootan & Leding, 2015).

Need for Cognition in Memory Research

Although NFC has its roots in judgment and decision-making research, this individual difference has also been explored in memory research. As such, several studies have shown that high-NFC and low-NFC individuals also differ in recall task performance. For instance, people high in NFC have been found to recall more information from argumentative texts (Cacioppo et al., 1983). Namely, Cacioppo et al. (1983) had participants read 300-word editorials that contained either strong or weak arguments about exam regulations. Regardless of the arguments' strength, participants who scored high in NFC recalled significantly more of the arguments than their low-NFC counterparts. Similarly, Kardash and Noel (2000) found that individuals who were high in NFC recalled more pieces of text compared to their low-NFC counterparts.

This finding that high NFC is related to better recall has also been extended to materials beyond editorials. For instance, both Peltier and Schibrowsky (1994) and Kuo et al. (2012) found that individuals high in NFC demonstrated better recall for advertising materials when compared to low-NFC counterparts. Moreover, Reid et al. (1995) found that those who scored high in NFC demonstrated both greater reading comprehension and recall for topics within health-related pamphlets.

However, a distinct area of memory research concerning NFC involves the Deese-Roediger-McDermott (DRM) paradigm (Graham, 2006; Leding, 2011; Wootan & Leding, 2015). The DRM paradigm is a false memory task where participants are asked to study a series of semantically related target words (e.g., *nurse, patient, hospital*). These target words are associatively related to a critical lure that is never shown during the study phase (e.g., *doctor*). When asked to report all of the studied words, participants are susceptible to falsely reporting that they remember studying the critical lure (Roediger & McDermott, 1995). One explanation behind why these false memories occur is the activation-monitoring account – which posits that both spreading activation within semantic memory networks and elaborative encoding facilitate false memories of the critical lure (Gallo & Roediger, 2002). Namely, the activation-monitoring account suggests that while encoding the target words, participants engage in both conscious and unconscious processing of semantically related items. Since the critical lure is a word that is highly related to the target words, associations to the critical lure are likely to be activated while the participants are studying the targets, contributing to false memories of the critical lure later. This can be seen in manipulations of DRM task instructions – studies that encourage more elaborative semantic processing of DRM lists have demonstrated increased false memory for critical lures (Chan et al., 2005; Thapar & McDermott, 2001; Toglia et al., 1999).

Studies looking at the relationship between NFC and DRM performance suggest that individuals high in NFC engage more in this sort of elaborative processing compared to their low-NFC counterparts. Specifically, individuals high in NFC may be more likely to think about the semantic associations related to each target during encoding, compared to low-NFC participants. For instance, Graham (2006) first investigated the role of NFC in the DRM paradigm in two experiments. The author first found that those with high NFC falsely recognized more critical lures than those classified as having low NFC. However, there was not a significant relationship between NFC and true recognition of the target words. Graham (2006) then replicated this finding in a second experiment.

Leding (2011) expanded upon Graham's (2006) work by showing that high-NFC participants demonstrate higher rates of both true and false recall, compared to their low-NFC counterparts. Specifically, the authors manipulated attention as participants underwent the DRM paradigm, in addition to testing participants' memories multiple times. Namely, during the study phase of the DRM paradigm, participants in the divided attention condition were also given an auditory task where they had to keep track of how many odd numbers were presented three times in a row. Participants with high NFC falsely recalled the critical lures more often than the low-NFC participants, regardless of whether they were permitted to give their full attention during encoding. Moreover, they also found that high-NFC participants were able to correctly recall more target words than low-NFC participants when the participants were able to give their full attention to the DRM task. However, no significant difference in true recall was seen when the participants were distracted during encoding. Additionally, high-NFC participants gradually recalled more critical lures with subsequent testing, suggesting that with more opportunities to retrieve, high-NFC individuals also take more opportunities to engage in more elaborative processing.

Wootan and Leding (2015) also replicated this finding when manipulating levels of processing in a DRM task. Specifically, participants were asked to process some of the target words deeply (by rating these words on pleasantness) and other target words in a shallower manner (by indicating whether or not a given target word had the letter "e" in it). Similar to Leding (2011), Wootan and Leding (2015) also found higher true and false recall rates in high-NFC individuals compared to low-NFC individuals. Namely, false recall was significantly different between NFC levels when the target words were shallowly processed, but not when the words were deeply processed. Specifically, compared to the low-NFC group, the high-NFC group falsely remembered the critical lure more often when the target words were shallowly processed. However, the NFC groups did not differ in false recall when the target words were deeply processed. As such, Wootan and Leding (2015) posit that a deep processing strategy may be beneficial to low-NFC participants in the sense that it may enable the processing style of the low-NFC participants to match that of the high-NFC participants, resulting in similar rates of false recall for this condition. Additionally, although both NFC groups produced levels of processing effects, the high-NFC participants, in comparison with the low-NFC participants, still produced higher true recall rates for both the deeply processed and shallowly processed words. This suggests that high-NFC participants naturally process items in a deeper, more elaborative fashion regardless of experimenter-provided instructions.

As demonstrated in the aforementioned levels of processing research, both memory strategies and NFC levels vary in the degree of deep processing that they facilitate during study (Wootan & Leding, 2015). Moreover, given that NFC appears to influence the effectiveness of levels of processing manipulations, it may be possible that this individual difference also impacts the benefits of other memory strategies as well. Some robust memory encoding strategies include generation (Slamecka & Graf, 1982), enactment (Engelkamp, 1998), production (Macleod et al., 2010), and drawing (Wammes et al., 2016). In the lab, the effectiveness of these strategies is tested by comparing a strategy in question (e.g., generation) with a baseline encoding task (e.g., reading). These comparisons have been explored both within participants (i.e., having each participant study a mixture of items – half being studied with an encoding strategy and the others with a baseline task), as well as between subjects (i.e., having one group of participants study items with an encoding strategy and another group study with the baseline task).

In addition to within and between-subject designs, all of these tasks have been tested with recognition and recall – each resulting in differences regarding the magnitude and presence of each effect. Recognition involves presenting participants with a mixture of items that they had studied ("old" items or "targets") alongside items that were never shown during the study phase ("new" items or "lures"). In old-new recognition tests, participants are shown items individually and are tasked with indicating which test items are the ones that they had studied. In other words, participants identify which items they can recognize as "old". Conversely, recall tests ask participants to report which items they can remember *without* being presented with these study items again. In a cued recall test, participants may be given incomplete information (e.g., a letter cue or an associate of the studied items) in order to facilitate recall. Meanwhile, a free recall test provides no additional information to the participant at all. In general, participants tend to have higher recognition performance than recall performance (Watkins & Todres, 1978). As participants are shown old and new items in a recognition test, they have more external cues with which to retrieve the study items from memory. Meanwhile, recall tests provide far fewer cues, leading to a more difficult memory test. However, the degree to which the conditions of encoding match the conditions of the memory test also influences performance – with memory performance increasing as encoding and retrieval conditions coincide (transfer-appropriate processing; Baddeley et al., 2020; Hirshman & Bjork, 1988).

Overall, memory performance and effect sizes for many robust encoding strategies tend to be highest for within-subjects recognition designs and lowest for between-subjects free recall designs. (Fawcett et al., 2023; Fernandes et al., 2018; MacLeod & Bodner., 2017; McCurdy et al., 2020; Roberts et al., 2022). This suggests that relative processing during encoding plays a role in the effectiveness of these strategies (see also McDaniel & Bugg, 2008). Moreover, it highlights how retrieval methods (i.e., recognition vs. recall) interact with encoding strategies to influence memory performance. As the current thesis will focus on the mnemonic benefits of drawing, generation, and production in free recall, the following sections will outline these particular memory phenomena, provide an overview of their explanatory models, and summarize their relations with elaborative encoding.

The Drawing Effect

The drawing effect is the mnemonic benefit seen for drawn words compared to words that are studied using a baseline encoding task (Wammes et al., 2016). Specifically, the original drawing effect paradigm asked participants to study a list of concrete words while engaging in one of two different memory strategies. For half of the items, participants were cued to draw a single picture representing the given word. Then for the remaining half, participants were asked to write out the given word repeatedly. When memory for the studied words was tested, participants tended to remember more of the words that they had drawn, compared to the ones that they had written. As such, the difference in average memory performance between the drawn words and written words is what is referred to as the drawing effect (Wammes et al., 2016).

The drawing effect appears robust across a number of designs and conditions (Fernandes et al., 2018). For instance, although the benefits of drawing over writing were originally seen for within-subjects designs, Wammes et al. (2016) also found the effect in between-subjects experiments as well. However, the magnitude of the drawing effect is larger for within-subject designs. This memory benefit is also found both in free recall (Wammes et al., 2016) and recognition tests (Wammes et al., 2018). Moreover, it has been explored outside the lab, particularly in lecture halls (Wammes et al., 2016; Wammes et al., 2017). As such, the applicability of drawing as an educational strategy has also been explored. Specifically, the drawing effect has been demonstrated in memory for scientific definitions (Wammes et al., 2017) and abstract concepts (Roberts & Wammes, 2021). Adaptations of the classic paradigm have also extended beyond typical undergraduate student samples, with drawing being found to help children learn new concepts (Thiede et al., 2022) and to help older adults remember everyday events (Tran et al., 2023).

Why do we see a Drawing Effect?

The main explanatory framework behind the drawing effect is the integrated-components model (Fernandes et al., 2018). Namely, the act of drawing may help one's memory for studied items because this task facilitates elaborative, motoric, and pictorial (i.e., visual) processing. The combination of these three types of processing appears to result in more distinct memory traces, in comparison to less demanding encoding techniques. Specifically, not only does drawing require one to think elaboratively about the subject matter's representation, but it also requires one to engage in motor action to physically draw the material. Then, when one has finished drawing, they have the features of their picture to process as well.

Support for the integrated-components model can be seen in studies that compare drawing with encoding strategies that are thought to encourage only a portion of the model's components. For instance, Wammes et al. (2019) compared memory performances between drawing, tracing, viewing, and imagining study items. While drawing encompassed elaborative, motoric, and visual components, each of the other tasks only represented a portion of these three aspects. For instance, the instruction to trace the study items elicited the motoric and pictorial components. But because participants did not need to come up with the visual reference on their own, this tracing task was missing the elaborative component typically elicited in drawing. As such, drawing while studying was found to benefit memory more than these other encoding tasks, with Wammes et al. (2019) suggesting that the overall combination of integrated drawing components led to more distinctive memories of drawn items.

Moreover, the elaborative component of the drawing effect has been compared to the act of paraphrasing definitions (Wammes et al., 2017). Namely, Wammes et al. (2017) had participants either draw pictures representing a scientific definition or to paraphrase the target items into a more concise phrase. Wammes et al. (2017) found that participants remembered both the drawn items and the paraphrased items equally. As both encoding strategies performed similarly despite the additional pictorial and motoric aspects of drawing, it highlights the significance of elaboration in episodic encoding (Craik & Lockhart, 1992). As such, it appears that drawing produces mnemonic benefits due, in part, to facilitating elaborative thinking.

Elaborative Encoding in Drawing

Although the elaborative component of drawing has been demonstrated in comparisons with other memory strategies, support for this role of drawing can be seen in the costs of drawing as well. For instance, although drawing outperforms writing in terms of item memory, it appears to negatively impact sequence memory (Jonker et al., 2019). Jonker et al. (2019) suggest that this is due to the enhanced elaborative processing that drawing encourages. Specifically, because this increase in elaborative processing causes participants to focus on the semantic meaning of the drawn words, this strategy disrupts the processing of inter-item associations such as item order.

Additionally, when employed in the DRM paradigm, drawing also appears to increase false recognition relative to writing (Meade et al., 2020). This finding supports the notion that drawing encourages semantic activation during encoding, leading to accidental activations and retrievals of the critical lure.

The Generation Effect

The generation effect refers to the memory benefit seen for self-generated material over experimenter-provided information (Slamecka & Graf, 1978). In other words, participants tend to remember material better when they handle and create that information themselves, as opposed to simply reading items given to them by the experimenter. In most generation effect experiments, participants are asked to study a series of items, such as words (Slamecka & Graf, 1978), nonwords (Johns & Swanson, 1988; Nairne & Widner, 1987), or numbers (McNamara & Healy, 2000). In a typical within-subject design, participants are asked to self-generate half of the stimuli and to simply read the other half while studying. How participants self-generate varies across studies, but they all involve generating a target response according to given cues and rules. For instance, in their seminal paper, Slamecka and Graf (1978) cued their participants with a stimulus word and the first letter of what they should generate (e.g., "rapid-f" to generate the response "fast"). They then had their participants generate responses that were either associated with (e.g, "lamp-light"), within the same category (e.g, "ruby-diamond"), opposite from (e.g, "long-short"), a synonym of (e.g, "sea-ocean"), or that rhymed with (e.g, "save-cave"), the stimulus word. But regardless of generation rules, participants tend to remember more of the generated responses than the read words when tested on their memory for all of the studied words (Bertsch et al., 2007; McCurdy et al., 2020).

Beyond Slamecka and Graf's (1978) first explorations into generation as a memory strategy, there have been numerous adaptations to the kinds of cues used to elicit generation responses. For instance, studies have cued participants using incomplete sentences (Peynircioğlu & Mungun, 1993), scrambled anagrams (Gardiner et al., 1989), mathematical equations (McNamara & Healy, 2000), and word definitions (Hourihan & MacLeod, 2007). How memory is tested also varies – with the generation effect being seen in recognition tests (Begg et al., 1989; Hunt & McDaniel, 1993; Slamecka & Graf, 1978), as well as in cued (Hirsh & Bjork, 1988; Slamecka & Graf, 1978) and free recall (Hunt & McDaniel, 1993; Slamecka & Graf, 1978). Similarly, the generation effect also appears in both within-subject and between-subjects designs, although the effect size tends to be larger when tested within-subjects – suggesting that generation facilitates multiple processes when benefiting memory (Bertsch et al., 2007; McCurdy et al., 2020).

Why do we see a Generation Effect?

Multiple theories have been suggested to help explain the benefit of generation. Concerning item memory, the most prevalent ones are the semantic activation theory and the multifactor transfer-appropriate processing account (Hirshman & Bjork, 1988; McCurdy et al., 2020; McElroy & Slamecka, 1982). The semantic activation theory posits that generation aids memory because the act of self-generating involves searching through prior semantic knowledge to find what is associated with the given cues (i.e., deep processing). This semantic activation then strengthens the memory trace along the same lines that deep processing does – the consideration of different semantic associations during encoding then allows for more internal retrieval cues when memory is later tested. Support for the semantic activation theory can be seen in how the generation effect has been found for meaningful stimuli like words, but not for nonsensical stimuli such as non-words (McCurdy et al., 2020). But as opposed to considering just one factor, the multifactor transfer-appropriate processing account considers both deep processing and shallow processing when it comes to generation (Hirshman & Bjork, 1988; McCurdy et al., 2020; McElroy & Slamecka, 1982). Specifically, the multifactor theory suggests that generation enhances both item-specific processing as well as relational processing. The benefits that arise from these different types of processing are then elicited based on the type of memory test (i.e., transfer-appropriate processing). Specifically, it is thought that increased item-specific processing aids in discerning studied target items from foils in a recognition task. As such, support for the notion that generation enhances item-specific processing can be seen when the generation effect occurs in recognition tests (Begg et al., 1989; Hirshman & Bjork, 1988). Meanwhile, enhanced relational processing is considered to strengthen memories of the relationships between the cue and target, as well as the relationships between the entire study list. These stronger memories about inter-item relationships allows the recall of one target item to then act as a reminder to facilitate the recall of another target. As such, support in favour of amplified relational processing is demonstrated in generation effects in both cued and free recall (Hunt & McDaniel, 1993; Hirshman & Bjork, 1988).

Elaborative Encoding in Generation

As highlighted in these item memory theories, although generation involves semantic activation, it is also thought to encourage item-specific processing as well (Hirshman & Bjork, 1988). This can also be seen in how generation boosts memory performance while avoiding increases in false memory reports. For instance, Soraci et al. (2003) found that during DRM tasks, generation was found to increase true memory accuracy without encouraging false memory reports of critical lures. Namely, participants who self-generated during the encoding phase both correctly recalled and recognized more studied words than individuals who simply read during encoding. However, the rate of false memories for critical lures was equal between the read and self-generate conditions. As such, Soraci et al. (2003) posit that self-generation increases item distinctiveness, resulting in a more robust verbatim memory trace, rather than a more general gist trace across both true and false items.

The Production Effect

The production effect is the memory benefit experienced for produced material over baseline items (MacLeod et al., 2010). For instance, participants tend to remember words better when the items are read aloud, rather than studied silently. The classic production effect paradigm asks participants to study a series of words – reading half of the words aloud and reading the other half silently. In a later memory test, participants tend to remember more of the produced words in comparison to the non-produced words.

First explored by Hopkins and Edwards (1972), the production effect was formally coined as such decades later by MacLeod et al. (2010). Since then, the production effect has been explored across several production methods, study items, experimental designs, and memory tests. For instance, the utility of production has been seen beyond simply speaking aloud – with memory benefits observed for writing, typing, mouthing, and singing (Forrin et al., 2012; Jamieson & Spear, 2014; Quinlan et al., 2019), and even just imagining that you are typing the words (Jamieson & Spear, 2014). Additionally, the kinds of study items that benefit from production also vary, with the production effect being found for words (MacLeod et al., 2010), nonwords (MacLeod et al., 2010), word-word pairs (Putnam et al., 2014), and word-picture pairs (Fawcett et al., 2012; Hourihan & Churchill, 2020). Moreover, while the production effect was initially defined as a purely within-subject phenomenon, meta-analytic investigations (Fawcett et al., 2013; Fawcett et al., 2023) have revealed that production tends to benefit memory in between-subjects designs as well – albeit to a lesser degree. Finally, the production effect has been seen in both within-subject tests of recognition and recall (MacLeod & Bodner., 2017), as well as in between-subject recognition tests (Fawcett et al., 2023). However, recall tests in between-subjects designs appear to show a more nuanced production effect, with production only benefiting items that are studied later in the encoding phase (Fawcett et al., 2023).

Why do we see a Production Effect?

An early explanation behind the mnemonic benefit of production is the distinctiveness heuristic (Hunt, 2013; MacLeod et al., 2010). Specifically, this account posits that production makes each produced item more distinctive than the silent items, as each of these items is processed with a unique piece of produced information at the time of encoding. This contrasts the silent items, which, relative to the produced items, become more indistinguishable from one another during encoding (Conway & Gathercole, 1987; MacLeod et al., 2010). At test, this relative distinctiveness then makes the memories for produced items easier to retrieve than the silent ones. In other words, participants remember more produced items because they can remember the specific act of producing that was associated with a given item. Support for this distinctiveness heuristic can be seen when produced items are made indistinct from one another. For instance, MacLeod et al. (2010) found that the production effect disappeared when participants were only instructed to say “yes” while producing, rather than read each produced item aloud.

Additionally, it is thought that speech in particular facilitates distinctiveness due to the translation of study items from a visual modality to an auditory one. For instance, Conway and Gathercole (1990) manipulated how participants studied a list of words. One group only engaged in visual studying – quietly reading half of the study items, while both reading and writing the other half. Another group of participants engaged in both visual and auditory studying – listening to half of the items, while both listening to and writing the remaining half. They found that participants remembered the most when they both listened and wrote – an act that engaged both visual and auditory modalities. This coincides with findings concerning the production effect, with the memory benefit of speech being stronger than the benefits of non-auditory strategies, such as writing, typing, and mouthing (Forrin et al., 2012). Speech involves processing both motoric and phonological information, which allows this particular mode of production to have more distinctive features than non-auditory strategies.

A key component of this account is that the difference in distinctiveness between produced and unproduced items is relative, with the idea that the effect of distinctiveness would disappear when participants are unable to compare the two conditions. But the benefit of production in between-subjects designs contradicts this notion of relative distinctiveness, as participants in these experiments would have no way of comparing aloud and silent items. As such, it is also thought that the act of production strengthens the memory trace, contributing to better memory for produced items without the need for comparison with non-produced items (i.e., the strength account; Bodner et al., 2014; Fawcett et al., 2023).

In addition to the distinctiveness heuristic and the strength account, computational models have outlined a newer explanation for the production effect (e.g., Caplan & Guitard, in press). Specifically, both the distinctiveness heuristic and the strength account have limitations. For the distinctiveness heuristic, the existence of between-subjects and pure-list production effects (Bodner et al., 2014; Fawcett et al., 2023) contradict the notion that production is only beneficial due to relative distinctiveness. Similarly, the strength account alone is unable to explain why participants can identify the source memory for produced and silent items, even when silent items are studied multiple times to increase the strengths of the associated memory traces (Ozubko et al., 2014).

As such, computational distinctiveness accounts present a combination of these original explanations. Like the distinctiveness heuristic, this modern distinctiveness account assumes that the act of producing enables more item-specific features to be encoded. But like the strength account, this contemporary framework also assumes that these produced features contribute to stronger memory traces. In Caplan and Guitard's (in press) computational model, they suggest certain features receive more attention based on encoding conditions. So, in the case of reading aloud, producing would encourage a participant to attend to more of the salient phonological features of an item than would reading silently. In particular, these phonological features from production are assumed to differ from those of semantic processing and imagery. While both produced and silent items are modelled to have an equal number of attended semantic features, participants focus more on the phonological features of the produced items.

Regardless, all attended features then become encoded into memory. To model recognition tests, the authors examined the degree to which the encoded features of the memory representation matched the features of a given test item. More of the produced items are recognized in these models because these produced items have more encoded features than their silent counterparts. In sum, computational models of the production effect help explain this phenomenon by considering a combination of distinctiveness and strength accounts.

Elaborative Encoding in Production

As the benefit of production is primarily explained by relative distinctiveness, there does not appear to be much support suggesting that the production effect is predominantly driven by semantic processing (but production does appear to encourage a degree of semantic processing – see Bailey et al., 2021; Fawcett et al., 2022; and Hourihan & Churchill, 2020; for support on the role of semantic processing in production). Rather, the benefit of production primarily appears because the act of production involves the additional encoding of item-specific features, increasing relative distinctiveness between produced and silent items. Support for the notion that the act of production encourages more item-specific processing can be inferred from mathematical models that suggest production encourages the strengthening of distinctive phonological features in memory (Caplan & Guitard, in press). Additionally, produced items still appear to be better remembered than non-produced items, even when participants also study the items using elaborative encoding, suggesting that production and elaborative encoding strategies at least operate in separate ways. For instance, both MacLeod et al. (2010) and Forrin et al. (2014) combined generation and production. Namely, MacLeod et al. (2010) asked participants to generate some words aloud while studying, while generating the remaining half silently. They found that the words that were generated aloud were still better remembered than the words that were generated silently – showing that production still benefits memory even when the episode is already elaboratively encoded by the act of generation.

Forrin et al. (2014) expanded upon Macleod et al. (2010) by conducting a more direct comparison between the generation effect and production effect. Specifically, they assessed the memory performance for words that participants had either generated aloud, generated silently, read aloud, or read silently. Although the generated items were better remembered than the read items, and the aloud items were better remembered than the silent items, Forrin et al. (2014) found no significant interaction between the generation and production conditions. Moreover, in a second experiment, Forrin et al. (2014) combined visual imagery and production. Participants either imagined words deeply (by imagining the physical representation of the word) or shallowly (by imagining that the word itself was in uppercase font). After each imagery instruction, participants read each word either aloud or silently. Like their first experiment, they found main effects of imagery and production, but also reported a lack of interaction between the two conditions. This lack of interaction in both experiments suggests that incorporating elaborative encoding does not significantly impact the magnitude of the production effect.

Present Research

While there has been research on NFC and memory, there have yet to be any investigations on how NFC may impact the relative memory benefit of specific encoding strategies like drawing, generation, and production. As individuals high in NFC have been shown to have greater recall than low-NFC counterparts (Leding, 2011; Wootan & Leding, 2015), it is possible that high-NFC participants would show greater overall memory performances in these drawing, generation, and production paradigms. Additionally, it is also possible that low-NFC individuals would have lower baseline memory performances than high-NFC counterparts. Such a relationship has been seen in other personality traits, such as Openness to Experience. For instance, older adults with high Openness scores also benefit less from generation than their less open counterparts. This smaller benefit was because more open individuals already possessed greater recall ability, so their baseline memory performances were already high (Taconnat et al., 2022). As such, it was expected that in the current experiments, low-NFC participants would benefit more from elaborative encoding strategies (e.g., drawing and generation) compared to the high-NFC participants.

Therefore, the current experiments sought to explore whether NFC impacts the magnitude of the drawing effect, generation effect, and the production effect. These replicated and extended the active encoding paradigms originally outlined by Wammes et al. (2016), Slamecka and Graf (1978), and MacLeod et al. (2010). Experiment 1 explored the drawing effect, Experiment 2 focused on the generation effect, and Experiment 3 investigated the production effect. All three experiments were within-subject designs and shared the same word lists and individual difference measures. As such, the experiments primarily differed in the paradigm. Experiment 1 was completed in-person, Experiment 2 was run online via Pavlovia and Qualtrics, and Experiment 3 was conducted both in-person and supervised online.

Each experiment had two predicted outcomes. First, it was expected that the drawing effect, generation effect, and production effect would be replicated. These memory effects have been frequently replicated in within-subjects designs (Fernandes et al., 2018; MacLeod & Bodner, 2017; McCurdy et al., 2020). Secondly, it was hypothesized that NFC level would show a negative relationship with the mnemonic benefit of elaborative encoding strategies. Specifically, high-NFC participants were expected to benefit less from more elaborative memory strategies when compared to their low-NFC counterparts. It was also hypothesized that this negative relationship would be most apparent for highly elaborative memory strategies. As such, the relationship between NFC and drawing was expected to be the most apparent, the relationship between NFC and benefit of generation was predicted to be weaker, and NFC was thought to bear little impact on the memory benefit of production.

These predictions are because NFC has been shown to have a positive relationship with elaborative processing (Leding, 2011; Wootan & Leding, 2015), a mental act that also improves memory (Craik & Lockhart, 1972). As the drawing effect and generation effect paradigms both provide external manipulations of elaborative encoding, it was predicted that high-NFC individuals, who are already internally motivated to think about the semantic meaning of the stimuli, would see less of a benefit to these strategies. This is because high-NFC individuals would already be engaging in elaborative processing for all of the stimuli, including the words within the baseline encoding condition. As such, the relative memory difference between the active encoding condition and the comparative condition would be smaller.

Namely, Experiment 1 compared memory performance between a “draw” condition and a baseline “write” condition to examine the magnitude of the drawing effect. In other words, during encoding, all participants were asked to draw half of the study items, while writing the remaining half. As drawing is considered to be a highly elaborative memory strategy (Fernandes et al., 2018; Jonker et al., 2019; Meade et al., 2020), it was expected that both high and low-NFC individuals would recall more drawn words than written words. However, because high-NFC participants are thought to be more inclined to voluntarily engage in elaborative processing for all of the words, then these individuals were also expected to recall more written words than their low-NFC counterparts. In other words, high-NFC participants were expected to have higher baseline memory performance than their low-NFC counterparts. Additionally, high-NFC participants were also hypothesized to recall a relatively similar amount of drawn words as the low-NFC participants. This is because the act of drawing would take up mental resources involved with elaborative processing. Therefore, even if high-NFC participants were more likely to engage with the semantic meaning of the drawn words on their own, they would already be doing so as a part of the task instructions. As such, the relative benefit would be different for each NFC level – low-NFC participants were expected to benefit more from drawing because they would be less likely to remember more written items on their own.

A similar set of predictions were hypothesized for Experiment 2. Specifically, memory performances between a “generate” and baseline “type” condition were compared. Namely, for stimuli in the "generate" condition, participants were shown definitions and letter cues representing to-be-remembered items. Participants were asked to generate what the items were based on the provided cues and to remember their guesses for later. For the "type" condition, participants were simply shown the item and asked to copy it in a textbox as they studied it for later. Similar to Experiment 1, it was expected that participants would recall more generated words than typed words, regardless of participants' NFC levels. However, high-NFC participants were also predicted to remember more typed words than their low-NFC counterparts due to naturally engaging in elaborative processing.

Then in Experiment 3, it was predicted that NFC would have no association with the magnitude of the memory benefit associated with production. The memory benefit of production was measured by comparing performance between items in "aloud" and "silent" conditions. Specifically, participants read half of the words aloud while studying, and read the remaining half silently. As producing is not known to involve semantic processing, it was thought that high-NFC participants would be able to spontaneously elaborate on top of producing during aloud trials – leading to improved memory relative to the low-NFC participants. However, these high-NFC participants were also expected to elaborate during silent trials, increasing recall performance for the baseline condition as well. As such, it was predicted that high-NFC and low-NFC participants would show different baseline performances but relatively equal performance differences between "aloud" and "silent" conditions (i.e., similar magnitudes of production effects).

In sum, it was predicted that low-NFC individuals would demonstrate a larger drawing effect and generation effect than their high-NFC counterparts. This is because these low-NFC participants would be expected to have a lower baseline performance in the comparative encoding conditions. Concerning the production effect, it was hypothesized that NFC would have little relation to the magnitude of this memory benefit as production does not appear to encourage semantic or elaborative processing.

Chapter 2: Experiment 1

In this first experiment, the primary aim was to determine whether differences in NFC level would be associated with differences in drawing effect magnitude. In other words, would NFC scores substantially predict the magnitude of the drawing effect? In line with previous drawing effect research, the present experiment compared two memory strategies – drawing and writing. If high-NFC individuals do engage in more elaborative encoding than low-NFC counterparts (Leding, 2011; Wootan & Leding, 2015), then it would be expected that high NFC scores would predict smaller differences in memory performance between the drawing and writing conditions. For instance, individuals high in NFC may be more inclined to connect written items with prior knowledge or to engage in imagery, in an attempt to memorize these items better. Meanwhile, low-NFC participants may be more inclined to simply follow the instructions about writing the items without partaking in any spontaneous elaboration. This would lead to high-NFC participants having higher recall for the written items in comparison to their low-NFC counterparts. This higher baseline would then result in a smaller relative memory benefit for drawing over writing in high-NFC participants.

As such, the first experiment employed a within-subjects drawing effect paradigm using free recall. Participants were shown a series of words and asked to study them. For a random half of the words, participants were asked to draw a picture representing the given word. For the remaining half of the words, participants were asked to simply write each word repeatedly until the end of the trial. After the encoding phase, participants were given a free recall test and measured on NFC, mental imagery ability, and demographic variables.

Method

Participants

A sample of 103 students was recruited from the Memorial University of Newfoundland in exchange for additional course credit. Participants were prohibited from signing up from more than one of the three experiments included in the current thesis. Two participants were excluded from the final data set due to experimental program crashes. Additionally, a portion of the participants experienced a coding error that resulted in an uneven distribution of items across encoding conditions (e.g., 12 “draw” items and 18 “write” items, rather than an equal number of 15 items in each condition). Participants who had within two trials of even distribution were retained. As such, 17 more participants were excluded, leaving a final data set of 84 participants.

The participants in the final data set ranged from 17 to 37 years of age (M = 21.18, SD = 3.12). Seventy-three of the participants self-identified as female, nine self-identified as male, and two self-identified as non-binary. Seventy-seven of the participants reported right-handedness and seven reported left-handedness. All participants reported fluency in English and gave informed consent to participate in the study. Data collection occurred during the Fall and Winter semesters of the 2023-2024 academic year (see Appendix A for ethics approval letter).

Materials

Encoding Stimuli. The stimuli were pulled from an 80-item target word list previously used by Wammes et al. (2016) (see Appendix B for stimuli list). These target words were the verbal labels of Snodgrass images (Snodgrass & Vanderwart, 1980), images of nouns considered to be visually simple and easy to draw. From these 80 targets, a 30-item subset was randomly selected for presentation during the encoding phase, such that each participant studied a unique list of items. Using PsychoPy 2023.2.3 (Peirce et al., 2019) on a computer running Windows 10 Enterprise, each word within the 30-item list was also randomly assigned to a draw or write condition. Stimuli were presented in the centre of a 17-inch Dell monitor screen in a black Open Sans font. Positioned above each target item was a text prompt indicating elaborative encoding instruction ("draw" vs. "write"). Participants were provided with 14 cm x 11 cm sheets of paper and a ballpoint pen.

Questionnaires. The 18-item Need for Cognition scale (NFC-18; Cacioppo et al., 1984) measures participants' tendency to exert effort when approaching cognitive problems. Participants are presented with 18 items describing various attitudes toward cognitive effort (e.g., "I prefer my life to be filled with puzzles I must solve") and rate their agreement with the item on a 9-point Likert scale (+4 = "Very strongly agree", -4 = "Very strongly disagree"). An additional item was added as an attention check ("Please answer 'Strongly disagree' for this question").

The Vividness and Visual Imagery Questionnaire (VVIQ; Marks, 1973) measures individual differences in mental imaging ability. Participants are asked to imagine different aspects of four separate scenes. Then, they rate on a 5-point Likert scale how vividly they can imagine these details (1 = "Perfectly clear and as vivid as normal vision"; 5 = "No image at all, you only 'know' that you are thinking of the object").

In addition to the NFC-18 and the VVIQ, participants were also asked demographic questions such as their gender identity, age, handedness, and fluency in English. They were also asked whether they had any previous drawing experience (see Appendix B).

Procedure

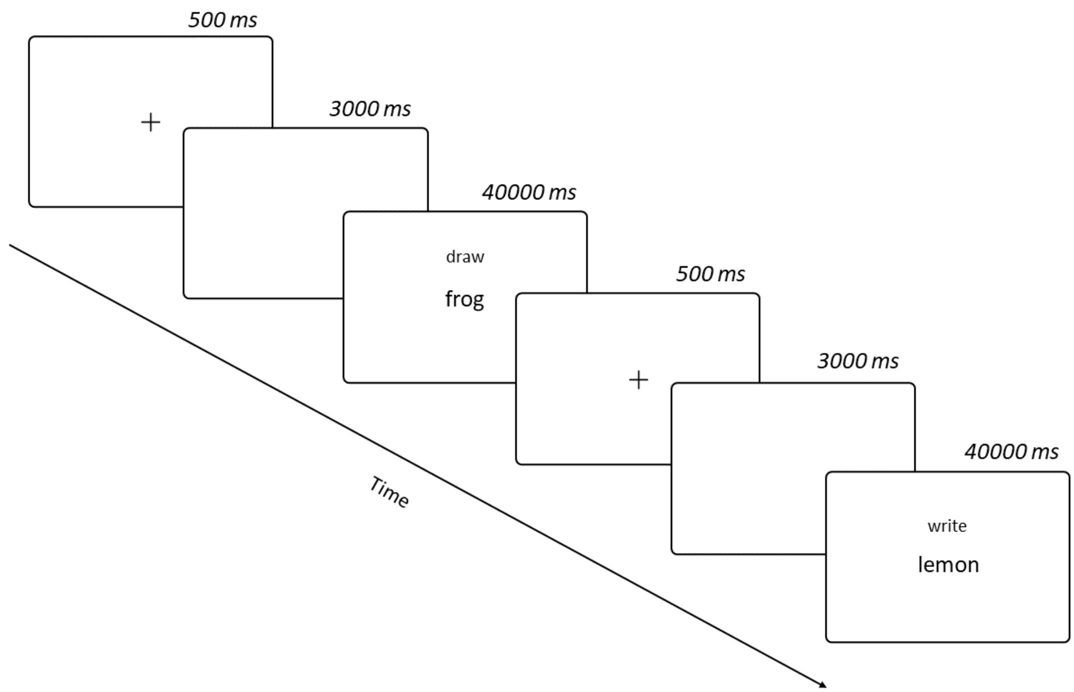
Participants were individually invited into a quiet testing room at the university's Metacognition Laboratory. Once seated, a researcher explained that the study would ask participants to memorize a series of words as they appeared onscreen one at a time. Participants were also told that as they studied each word, there would also be a prompt indicating whether to draw or write the study word. When the prompt said "draw", participants were asked to draw a single picture representing the target word and to continue adding detail to their image for the duration of the trial, when a tone would indicate to stop drawing. When the prompt said "write", participants were asked to write the study word repeatedly until a tone signalled the end of the trial. These instructions were also presented on the screen, and participants were given time to read them over.

Before beginning the experiment, the participants underwent two practice trials – one that asked them to draw the presented word and another that asked them to write out the presented word. When needed, participants were given additional instructions based on their practice performances, After completing the practice phase, participants were reminded that regardless of whether they write or draw the word, to do their best to remember all of the study items.

Encoding Phase. The encoding phase consisted of 30 trials. On each trial, a 500 ms fixation cross appeared. This was followed by a study word in the centre of the screen with an instruction prompt positioned above it, which both remained on the screen for 40 seconds. Participants had these 40 seconds to either draw or write the study item, according to the prompt. After 40 seconds, a 500 ms tone sounded, indicating that participants had to stop their elaborative task, flip over their paper, and get a new piece of paper for the next trial (see Figure 1).

Figure 1

Example of Encoding Trials for Experiment 1



Filler Task. A 30-second parity task was programmed between the study phase and test phase. This task presented a single, two-digit number in the centre of the screen. Participants were asked to indicate the parity of the number by pressing the "e" key when the number was even, and the "o" key when the number was odd. The trials would only change when a response was given, but participants were instructed to respond as quickly and as accurately as they could. The task continued to present new numbers until 30 seconds passed.

Recall Phase. Participants were presented with 30 textboxes within which they typed in their responses. The textboxes were arranged in a matrix of five rows and six columns. Participants could submit their answers after 60 seconds but were given as much time as desired to answer and modify their responses. All textboxes and inputted responses remained onscreen until participants submitted their final answers.

Questionnaires. Following the free recall test, the participants were provided with demographic questions, the NFC-18, and the VVIQ. A portion of the participants filled out the questionnaires outside of the laboratory while the rest completed the measures immediately after the recall test.

Results

Statistical Analyses

The data were cleaned using the *tidyverse* 2.0 package (Wickham et al., 2019) in R 4.3.1 (R Core Team, 2023). A combination of frequentist and Bayesian statistics was then run in jamovi 2.3.28 (The jamovi project, 2023). Frequentist analyses (including t-tests and linear regression) were calculated using the base *jmv* modules (The jamovi project, 2023), with the *moretests* modules (Love & Moreno, 2024) being used for additional assumptions testing. Meanwhile, the Bayesian linear regressions were calculated using the *jsq* module with default priors (Clyde et al., 2011; Clyde, 2017; JASP, 2018).

Recall Test Performance

Recall performance was calculated as the proportion of successfully recalled study words. Specifically, each study word that was assigned to each participant was coded as either a 1 (“recalled”) or 0 (“not recalled”). Scoring was liberal in the sense that plural items and synonyms were coded as correct. The proportion of recalled words was then averaged across participants and encoding conditions (“draw” vs. “write”). Overall, participants recalled $M = 0.52$ ($SD = 0.14$) of the words.

To determine if drawing was a more effective memory strategy than writing, a paired-samples t-test was performed on the proportion of correctly recalled items, with encoding condition as a within-subjects factor. A nonsignificant Shapiro-Wilks test, Kolmogorov-Smirnov test, Anderson-Darling test (all $p > .05$), and a visual inspection of the associated Q-Q plot all suggested that the data met the assumption of normality. As such, a Student's paired samples t-test was used. There was a significant difference between encoding conditions, with participants recalling significantly more drawn words ($M = 0.64$, $SD = 0.19$) than written words ($M = 0.41$, $SD = 0.16$), t (83) = 10.3, $p < 0.01$, $d = 1.12$.

Influence of NFC on Memory Performance

Descriptive statistics regarding NFC-18 and VVIQ performance can be found in Table 1.

Table 1

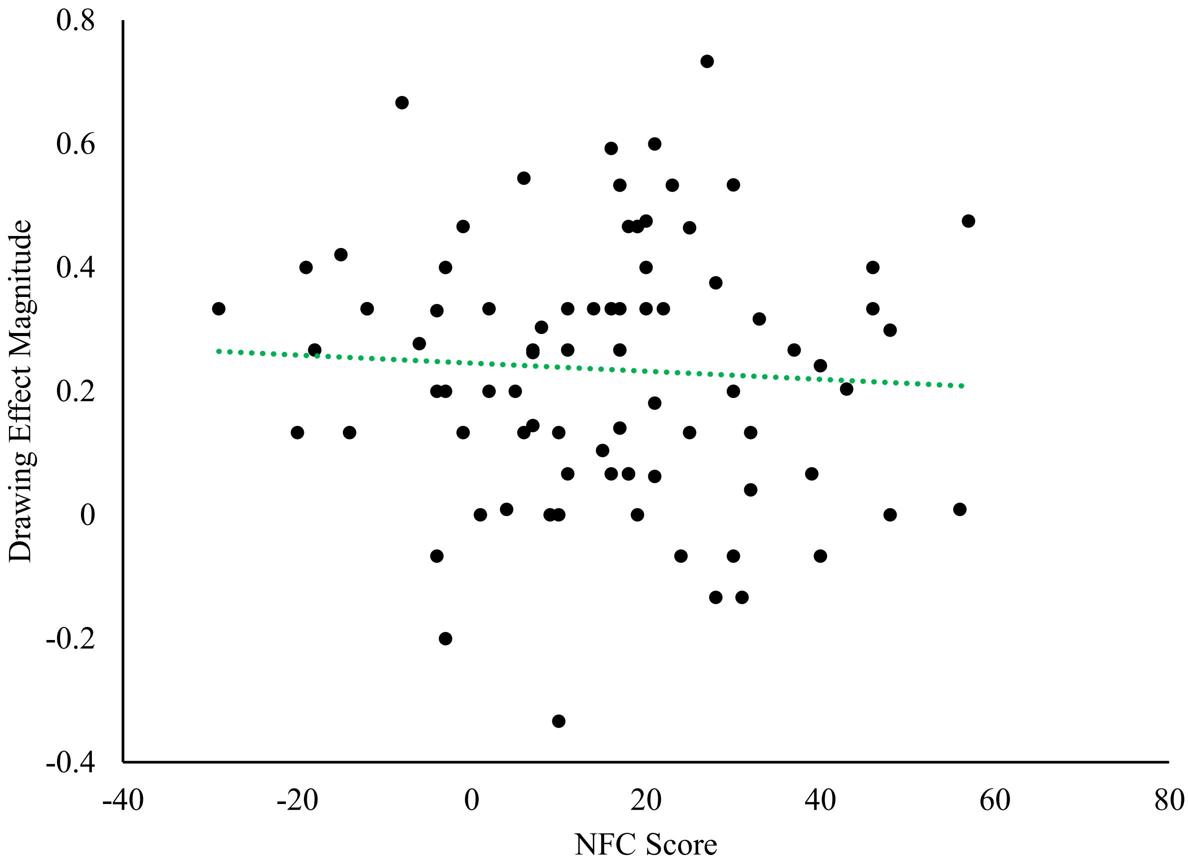
Descriptive Statistics for NFC-18 and VVIQ Scores

Experiment	N	Scale	Mean	SD	Median	Range
1	89	NFC-18	15.2	18.1	17.0	-29-57
		VVIQ	62.3	10.4	62.5	28-80
2	102	NFC-18	11.1	16.9	13.0	-41-49
		VVIQ	59.9	12.3	62.0	16-80
3	32	NFC-18	10.4	15.1	8.5	-19-50
		VVIQ	56.8	12.2	55.5	16-80

To determine if NFC had an influence on the magnitude of the drawing effect, a linear regression was performed between the NFC scores and the difference scores between the encoding conditions. NFC score was not a significant predictor of the drawing effect's magnitude, F (1, 82) = 0.29, p = .589, R = .059, R^2 = .004, Adjusted R^2 = -.009 (see Figure 2). Concerning recall performance in the active and baseline encoding conditions separately, NFC did not significantly predict the mean number of drawn words that were recalled, F (1, 82) = 0.92, p = .341, R = .105, R^2 = .011, Adjusted R^2 = -.001. Moreover, NFC did not significantly predict the mean number of written words that were recalled, F (1, 82) = 3.37, p = .07, R = .199, R^2 = .039, Adjusted R^2 = .003. NFC score was also not a significant predictor of overall recall performance, F (1, 82) = 2.49, p = .118, R = .172, R^2 = .027, Adjusted R^2 = .018.

Figure 2

Scatterplot of the Relationship Between Drawing Effect Magnitude and NFC Score



As these nonsignificant findings are inconclusive, corresponding Bayesian analyses were also run to determine how much support could be provided to the null and alternative hypotheses. A Bayesian linear regression was performed using default priors to ascertain whether NFC substantially influenced the difference scores between the draw and write conditions. This analysis suggested that the null hypothesis is 3.86 times more likely than the alternative hypothesis, demonstrating moderate evidence in favour of NFC having no substantial effect on the magnitude of the drawing effect (BF_{10} = 0.259). With the active and passive encoding conditions, a Bayesian linear regression suggested that, given the current data, NFC has an inconclusive relationship with the average number of drawn words that were recalled.

Specifically, the null hypothesis appeared to be 2.95 times more likely than the alternative hypothesis (BF_{10} = 0.339). However, the relationship between NFC and recall performance for the written words appeared inconclusive – with a Bayesian linear regression only suggesting that the null hypothesis was only 1.02 times more likely than the alternative hypothesis (BF_{10} = 0.976). Concerning NFC and overall recall, a Bayesian linear regression with default priors suggested that the null hypothesis is only 1.49 times more likely than the alternative hypothesis (BF_{10} = 0.673). As such, given the current data surrounding NFC and overall recall performance, the evidence in favour of either hypothesis remains inconclusive.

Exploratory Analyses

As Wammes et al. (2016) tested whether scores on the VVIQ were associated with memory performance, exploratory analyses involving VVIQ were also conducted for the present experiments, VVIQ score was not significantly predictive of the drawing effect's magnitude, F (1, 82) = 2.21, p = .141, R = .162, R^2 = .026, Adjusted R^2 = .001, nor was the scale a significant predictor of overall recall performance, F (1, 82) = 0.30, p = .585, R = .0605, R^2 = .0004, Adjusted R^2 = -.0085. However, NFC scores were a significant predictor of VVIQ scores, such that NFC scores accounted for approximately 8.01% of the variance in VVIQ scores, F (1, 82) = 8.11, p = .005, R = .299, R^2 = .090, Adjusted R^2 = .008. This finding is supported by the Bayes inclusion factor from a corresponding Bayesian linear regression, which indicated that the alternative hypothesis is 7.14 times more likely than the null hypothesis (BF_{10} = 7.14).

Of note, although the NFC-18 used in the current experiments was measured on a 9-point Likert scale, a 5-point scale is commonly used in the literature (Cacioppo et al., 1996). As such, the NFC data were also collapsed into a 5-point scale. Descriptive statistics involving these collapsed data can be found in Table 2.

Table 2

Descriptive Statistics for Collapsed NFC-18 Scores

Experiment	Mean	SD	Median	Range
1	62.1	9.42	63.0	40-81
2	60.8	9.35	61.5	29-78
3	60.3	8.64	59.0	45-80

Discussion

In Experiment 1, participants drew and wrote while encoding a series of words for a later recall test. Overall, this experiment found that participants recalled more drawn words than written words. As such, the expected within-subjects drawing effect was replicated in free recall. However, the magnitude of this drawing effect did not appear to have a significant relationship with NFC. Specifically, NFC did not significantly predict performance in either the "draw" or "write" conditions – a finding that contradicts the established hypotheses. Additionally, there did not appear to be a significant relationship between participants' NFC score and their overall recall performance, contrasting what has been previously found in the literature. However, Bayesian analyses suggested that this may not necessarily mean a complete lack of relationship between NFC and recall. Rather, evidence indicating whether this relationship exists or not remains inconclusive.

Surprisingly, although VVIQ did not significantly predict memory performance, a relationship did appear between VVIQ and NFC. In particular, higher NFC scores were associated with higher VVIQ scores. As the NFC-18 and VVIQ have not been directly compared in prior research, it is not immediately evident how this finding can be interpreted. As prior research has suggested that high-NFC individuals prefer to process verbal information over visual information (Sojka & Giese, 2001; Venkatraman et al., 1990), the observed pattern between NFC and VVIQ may be due to demand characteristics. Namely, because participants were first asked to draw before filling out any of the questionnaires, high-NFC participants may have felt obligated to provide higher ratings on the VVIQ.

Chapter 3: Experiment 2

As the first experiment focused on the drawing effect, the second experiment moved to generation – a memory strategy known to improve both elaborative and item-specific processing. Although traditional generation effect paradigms compare memory performance after generating and reading, the present experiment compared generating and typing in an online format. As participants were unsupervised in this design, the decision to use typing rather than reading was made to check whether participants were paying attention to the baseline items, rather than simply skipping over them. In line with the hypothesis for Experiment 1, it was expected that higher NFC scores would predict smaller differences in memory performance between the generating and typing conditions. For instance, high-NFC participants may be more inclined to think about the semantic meaning of typed items in order to improve memory for these items. Meanwhile, low-NFC participants may be more inclined to simply follow the instructions about typing out the baseline items without attempting to process them deeply. This would lead to high-NFC participants having higher recall for the typed items in comparison to their low-NFC counterparts. This higher baseline would then result in a smaller relative memory benefit for generating over typing in high-NFC participants.

As such, the second experiment employed a modified within-subjects generation effect paradigm. For half of the trials, participants were asked to generate and memorize words in accordance with given cues. For the remaining half of the trials, participants were shown words to memorize and asked to type them out as they studied them. After the encoding phase, participants were given a free recall test and measured on NFC, mental imagery ability, and demographic variables.

Method

Participants

A separate sample of 174 students from the Memorial University of Newfoundland were recruited online in exchange for additional course credit. Of the original 174 sign-ups, 69 withdrew from the study. The remaining 105 students completed the consent form and provided complete data. An additional three participants were excluded due to errors connecting their memory performance data with their questionnaire data. As such, the final analysis sample consisted of 102 participants. These participants ranged from 17 to 42 years of age (M = 20.9, SD = 3.58). Seventy-eight of the participants self-identified as female and 24 self-identified as male. Ninety-five participants reported right-handedness, six reported left-handedness, and one reported being ambidextrous. All participants in the final sample reported fluency in English and gave informed consent to participate in the study. Data collection occurred during the Fall and Winter semesters of the 2023-2024 academic year.

Materials

Encoding Stimuli. The study list was comprised of the same 80-item word list described in Experiment 1. However, instead of “draw” and “write” prompts, definitions were created for each word to serve as generation cues (e.g., “a green amphibian that hops around and eats flies.

f…" to cue the word "frog"; see Appendix B). Out of the original 80 items, 30 were randomly selected for study such that each participant studied a unique list of words. Half of the target items were then randomly selected to be displayed as words while the other half were randomly selected to be presented as generation cues. Stimuli presentation was programmed using PsychoPy 2023.1.3 (Peirce et al. 2019) and hosted online using Pavlovia (https://pavlovia.org/). All stimuli were displayed in black font, however, font types were converted to the default fonts of the participants' operating systems.

Questionnaires. The questionnaires were the same as those used in Experiment 1.

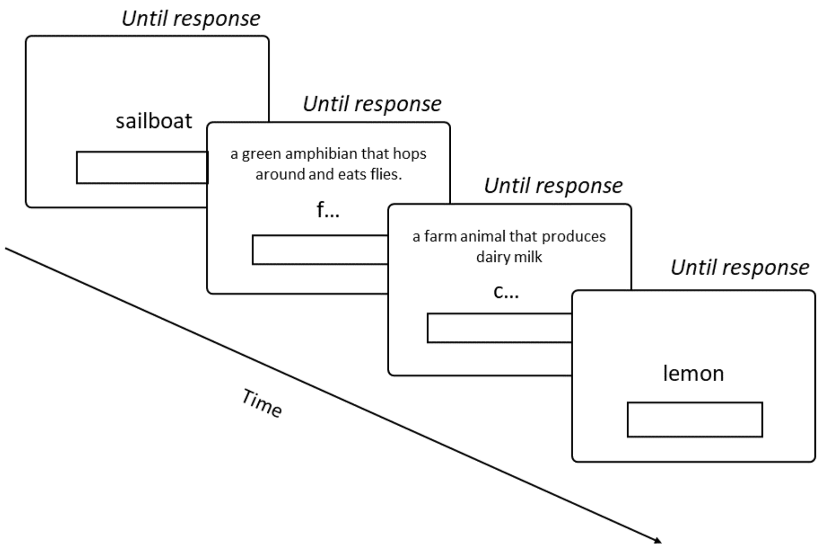
Procedure

Participants were tested online using Qualtrics and Pavlovia. They were first recruited online using Memorial University of Newfoundland's Sona system. From here, participants were provided with a link to a Qualtrics consent form. After completing the consent form, participants were redirected to the experiment hosted on Pavlovia. Instructions indicated that participants would be presented with a mixture of words – for some, participants would be asked to copy a displayed word (type condition) while for others, participants were asked to guess a word based on a given clue (generation condition). Participants were instructed to memorize all the words they typed out, regardless of whether they copied or guessed the word. Two practice trials were run prior to beginning the experiment – one trial representing a read trial and the other representing a generate trial.

Encoding Phase. The encoding phase consisted of 30 trials. On each trial, a 500 ms fixation cross appeared, followed by a 3000 ms blank screen. This was followed by a study trial, which differed based on condition. For words in the type condition, a study word was presented in the centre of the screen, with a textbox underneath. For words in the generate condition, a definition appeared at the top of the screen, the first letter of the target item was presented in the centre of the screen, and a textbox was positioned at the bottom of the screen. When participants finished either copying or guessing the target item, they pressed the "enter" key, which prompted the next trial to appear (see Figure 3).

Figure 3

Example of Encoding Trials for Experiment 2



The filler task and free recall test were identical to those described in Experiment 1. Following completion of the generation experiment, participants were provided with a code to enter into their Qualtrics consent forms. Upon inputting this code, the consent form unlocked access to the same questionnaires outlined in Experiment 1.

Results

Encoding Interval Analyses

As Experiment 2 was self-paced, the average amount of time that participants spent encoding was calculated. Overall, the encoding intervals were M = 5.86 (SD = 4.16) seconds long. Moreover, the average amount of time spent for each encoding condition (“generate” vs. “type”) was compared. A significant Shapiro-Wilks test, Kolmogorov-Smirnov test, and Anderson-Darling test suggested a violation of the normality assumption. As such, a Wilcoxon’s rank paired t-test was performed. The encoding intervals did significantly differ between generate and type conditions, with participants spending more time generating (M = 8.62, SD = 6.97) than typing (M = 3.09, SD = 2.71), W = 5251.0, $p < 0.01$, $r_{rs} = 0.99$.

Recall Test Performance

The data for Experiment 2 were cleaned and analysed following the same process as Experiment 1. Like in Experiment 1, recall performance was calculated as the proportion of successfully recalled study words. The proportion of recalled words was then averaged across participants and encoding conditions (“generate” vs. “type”). Overall, participants recalled M = 0.397 (SD = 0.17) of the words.

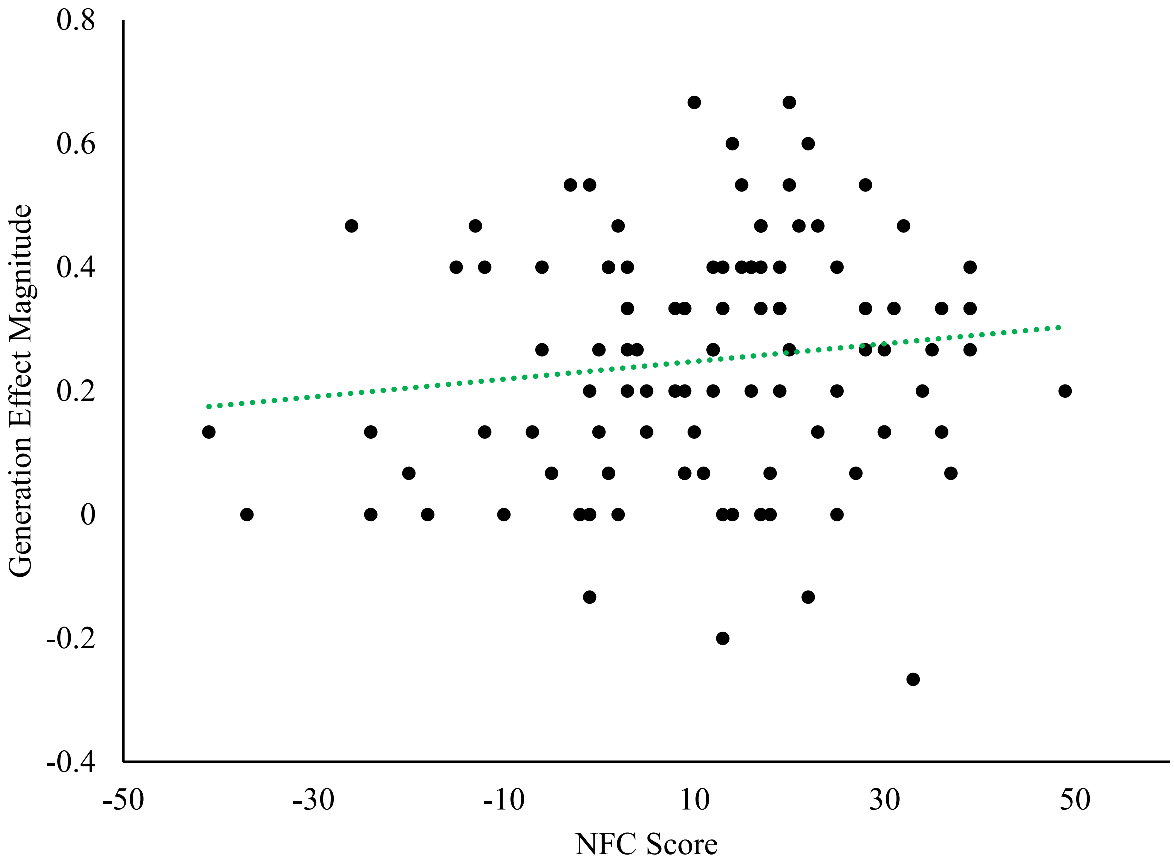
To determine if generation was a more effective memory strategy than typing, a paired-samples t-test was performed on the proportion of correctly recalled items, with encoding condition as a within-subjects factor. A nonsignificant Shapiro-Wilks test, Kolmogorov-Smirnov test (both $p > .05$), and a visual inspection of the associated Q-Q plot all suggested that the data met the assumption of normality. However, a significant Anderson-Darling test suggested non-normality. As such, a nonparametric Wilcoxon’s rank paired t-test was used. There was a significant difference between encoding conditions, with participants recalling significantly more generated words (M = 0.519, SD = 0.172) than typed words (M = 0.27, SD = 0.22), W = 4001.5, $p < 0.01$, $r_{rs} = 0.95$.

Influence of NFC on Memory Performance

As with Experiment 1, descriptive statistics regarding NFC-18 and VVIQ performance can be found in Table 1. To determine if NFC had an influence on the magnitude of the generation effect, a linear regression was performed between the NFC scores and the difference scores between the encoding conditions. NFC score was not a significant predictor of the generation effect's magnitude, F (1, 100) = 1.59, p = .209, R = .125, R^2 = .002, Adjusted R^2 = .006 (see Figure 4). Concerning the active and baseline encoding conditions, NFC did not significantly predict the mean number of generated words that were recalled, F (1, 100) = 0.207, p = .649, R = .046, R^2 = .001, Adjusted R^2 = -.008. Moreover, NFC did not significantly predict the mean number of typed words that were recalled, F (1, 100) = 2.18, p = .143, R = .146, R^2 = .021, Adjusted R^2 = .001. Finally, NFC score was also not a significant predictor of overall recall performance, F (1, 100) = 1.36, p = .246, R = .116, R^2 = .013, Adjusted R^2 = .004.

Figure 4

Scatterplot of the Relationship Between Generation Effect Magnitude and NFC Score



As these nonsignificant findings are inconclusive, corresponding Bayesian analyses were also run to determine how much support could be provided to the null and alternative hypotheses. A Bayesian linear regression was performed using default priors to ascertain whether NFC substantially influenced the difference scores between the generate and type conditions. This analysis suggested that the null hypothesis is only 2.36 times more likely than the alternative hypothesis, demonstrating inconclusive evidence in favour of NFC having no substantial effect on the magnitude of the generation effect (BF_{10} = 0.424). Similarly, the relationship between NFC and recall performance for the typed words appeared inconclusive – with a Bayesian linear regression only suggesting that the null hypothesis was only 1.83 times more likely than the alternative hypothesis ($BF_{10} = 0.548$). However, a Bayesian linear regression found moderate evidence suggesting that NFC has no relationship with the average number of generated words that were recalled. Specifically, the null hypothesis appeared to be 4.38 times more likely than the alternative hypothesis ($BF_{10} = 0.229$). Concerning NFC and overall recall, a Bayesian linear regression with default priors suggested that the null hypothesis is only 2.60 times more likely than the alternative hypothesis ($BF_{10} = 0.384$). As such, given the current data surrounding NFC and overall recall performance, the evidence in favour of either hypothesis remains inconclusive.

Exploratory Analyses

VVIQ score was significantly predictive of the generation effect's magnitude, F (1, 100) = 7.85, p = .006, R = .269, R^2 = .0728, Adjusted R^2 = .0635. This finding is supported by a corresponding Bayesian linear regression, which indicated that the alternative hypothesis is 6.38 times more likely than the null hypothesis ($BF_{10} = 6.38$). However the scale was not a significant predictor of overall recall performance, F (1, 100) = 1.65, p = .202, R = .128, R^2 = .016, Adjusted R^2 = .0064. NFC scores were a significant predictor of VVIQ scores, such that NFC scores accounted for approximately 5.59% of the variance in VVIQ scores, F (1, 99) = 6.37, p = .001, R = .245, R^2 = .059, Adjusted R^2 = .051. This finding is supported by the Bayes inclusion factor from a corresponding Bayesian linear regression, which indicated that the alternative hypothesis is 3.40 times more likely than the null hypothesis ($BF_{10} = 3.40$).

As in Experiment 1, the NFC-18 used in Experiment 2 was measured on a 9-point Likert scale. For ease of comparison with previous research, the NFC data were also collapsed into a 5-point scale. Descriptive statistics involving these collapsed data can be found in Table 2.

Discussion

In this experiment, participants generated and typed a series of words in preparation for a recall test. Participants recalled more generated words than written words, replicating the generation effect in a within-subjects, free recall design. But similar to Experiment 1, the magnitude of the generation effect did not appear to have a significant relationship with NFC. Specifically, NFC did not significantly predict how many typed items were recalled, nor did NFC have a significant relationship with the proportion of generated items that were recalled. But unlike Experiment 1, this lack of relationship between NFC and memory benefit was not supported by Bayesian analyses – which suggested inconclusive evidence given the current data.

Additionally, participants' NFC score did not significantly predict their overall recall performance. As discussed in Experiment 1, this lack of relationship between NFC and recall contradicts what has been found in the literature. However, Bayesian analyses also suggested that this finding is indeterminate given the current data.

Similar to Experiment 1, VVIQ did not significantly predict memory performance. But unlike the first experiment, VVIQ and NFC appeared to have a significant negative relationship. Namely, higher NFC scores were associated with lower VVIQ scores. While it has been found that high-NFC individuals prefer to process verbal information over visual information (Sojka & Giese, 2001; Venkatraman et al., 1990), this contrast with Experiment 1 is discussed further in the General Discussion.

In comparison with the first experiment, the recall performance for generated words in Experiment 2 was lower than the number of drawn words recalled in Experiment 1. This supports the multimodal encoding model, which suggests that drawing is an effective memory strategy because it facilitates elaborative, motoric, and visual processing – the combination of which provides more internal cues for ease of later retrieval. As the generation task used in this experiment primarily addresses elaborative processing, the smaller memory benefit could be due to generation involving fewer actions than drawing. However, this smaller memory benefit could also be attributed to other design differences between the two experiments, such as Experiment 1 presenting longer encoding intervals and Experiment 2 comprising of an online format.

Chapter 4: Experiment 3

The third experiment explored the relationship between NFC level and the production effect. As such, two encoding conditions were compared – reading aloud and reading silently. It was predicted that high-NFC participants would not significantly differ from low-NFC participants in terms of production benefit. This is because the production effect is not considered to be primarily driven by semantic processing (e.g., Forrin et al., 2014; MacLeod et al., 2010; but see Bailey et al., 2021; Fawcett et al., 2022; and Hourihan & Churchill, 2020; for support on the role of semantic processing in production). Therefore, participants could still engage in elaborative processing for both the aloud *and* silent items. As such, it was expected that regardless of NFC level, all participants would recall more words that were read aloud than words than were read silently. Additionally, the recall performance for silently studied words was expected to still be higher for high-NFC participants. However, the recall performance for aloud words was expected to be higher as well. As such, both silent and aloud condition performances were expected to be higher for high-NFC participants. However, the relative difference in memory performance between the aloud and silent conditions was not expected to significantly differ between high-NFC and low-NFC.

Therefore, the third experiment featured a within-subjects production effect paradigm. During the encoding phase, participants were asked to study a series of words. For half of the trials, participants were asked to read the given word aloud. Then for the remaining half of the trials, participants were asked to study the word silently. After the encoding phase, participants were given a free recall test and measured on NFC, mental imagery ability, and demographic variables.

Method

Participants

An additional 41 students from the Memorial University of Newfoundland participated in exchange for supplementary course credit. Individuals who had participated in Experiment 1 or Experiment 2 were ineligible for sign-up. Similar to Experiment 1 and 2, the target sample size was 100 participants. However, although data collection took place alongside the first two experiments (during the Fall and Winter semesters of the 2023-2024 academic year), fewer participants signed up for this study, resulting in an incomplete sample. As such, this thesis presents the current findings based on the sample to date. Similar to Experiment 1, a coding error resulted in a number of participants experiencing an uneven distribution of items across encoding conditions. Participants who had within two trials of even distribution were retained. As such, 9 participants were excluded, leaving a final data set of 32 participants.

These participants ranged from 18 to 33 years of age ($M = 21.7$, $SD = 3.63$). Twenty-four of the participants self-identified as female and eight self-identified as male. Twenty-nine self-reported right-handedness and three reported left-handedness. All participants reported fluency in English and gave informed consent to participate in the study.

Recruitment for Experiment 3 was conducted both in-person and supervised online. Of the 41 total participants, 34 completed the experiment while in the lab with the researcher present and seven participated online. The online participants were supervised by the lead researcher over Webex.

Materials

The study items for Experiment 3 consisted of the same list as Experiment 1 and were presented using PsychoPy 2023.2.3 (Pierce et al., 2019). Each word was positioned in the centre of a 17-inch Dell monitor screen in Open Sans font on a white background. Font colour indicated production instruction such that when the font colour was blue, participants were asked to read the word aloud and when the font colour was orange, participants were asked to study the word silently.

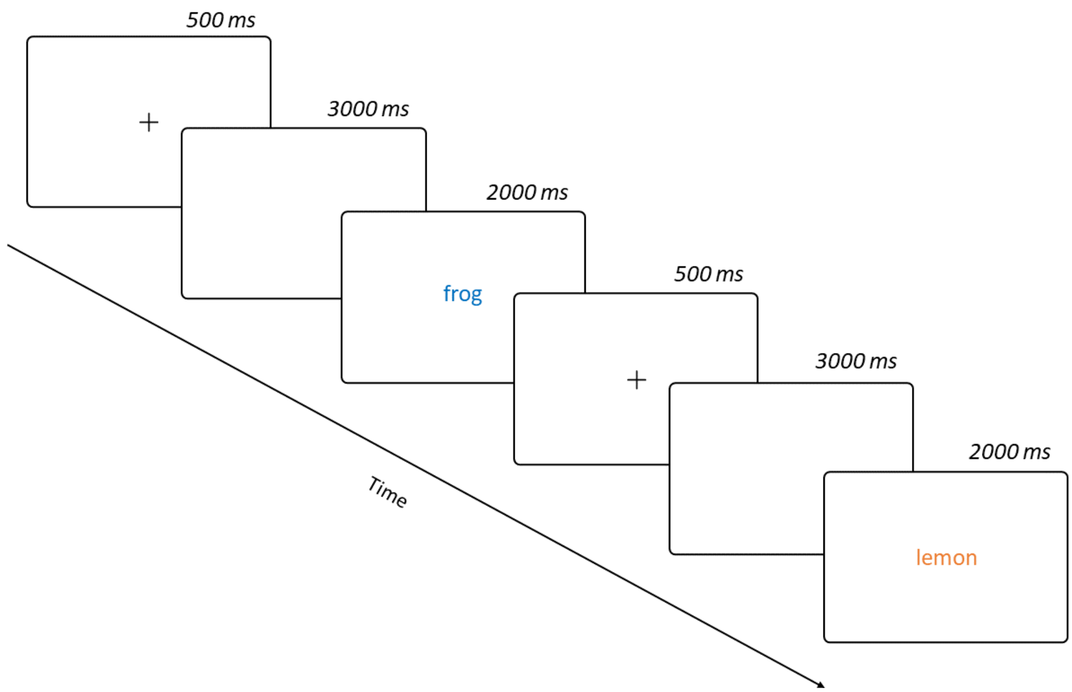
Procedure

Similar to Experiment 1, the encoding phase consisted of 30 trials. Each trial featured a 500 ms fixation cross, followed by a study word in the centre of the screen. This study word remained on the screen for 2 seconds. During this time, participants read blue words aloud and studied orange words silently (see Figure 5).

The filler task, free recall test, and questionnaire administration were identical to those outlined in Experiment 1. However, in Experiment 3, all of the participants filled out the questionnaires immediately after completing the free recall test.

Figure 5

Example of Encoding Trials for Experiment 3



Results

Recall Test Performance

The data for Experiment 3 were cleaned and analysed using similar methods to Experiments 1 and 2. Like in Experiment 1, recall performance was calculated as the proportion of successfully recalled study words. The proportion of recalled words was then averaged across participants and encoding conditions (“aloud” vs. “silent”). Overall, participants recalled M = 0.26 (SD = 0.089) of the words.

To determine if production was an effective memory strategy, a paired-samples t-test was performed on the proportion of correctly recalled items, with encoding condition as a within-subjects factor. A nonsignificant Shapiro-Wilks test, Kolmogorov-Smirnov test, and Anderson-

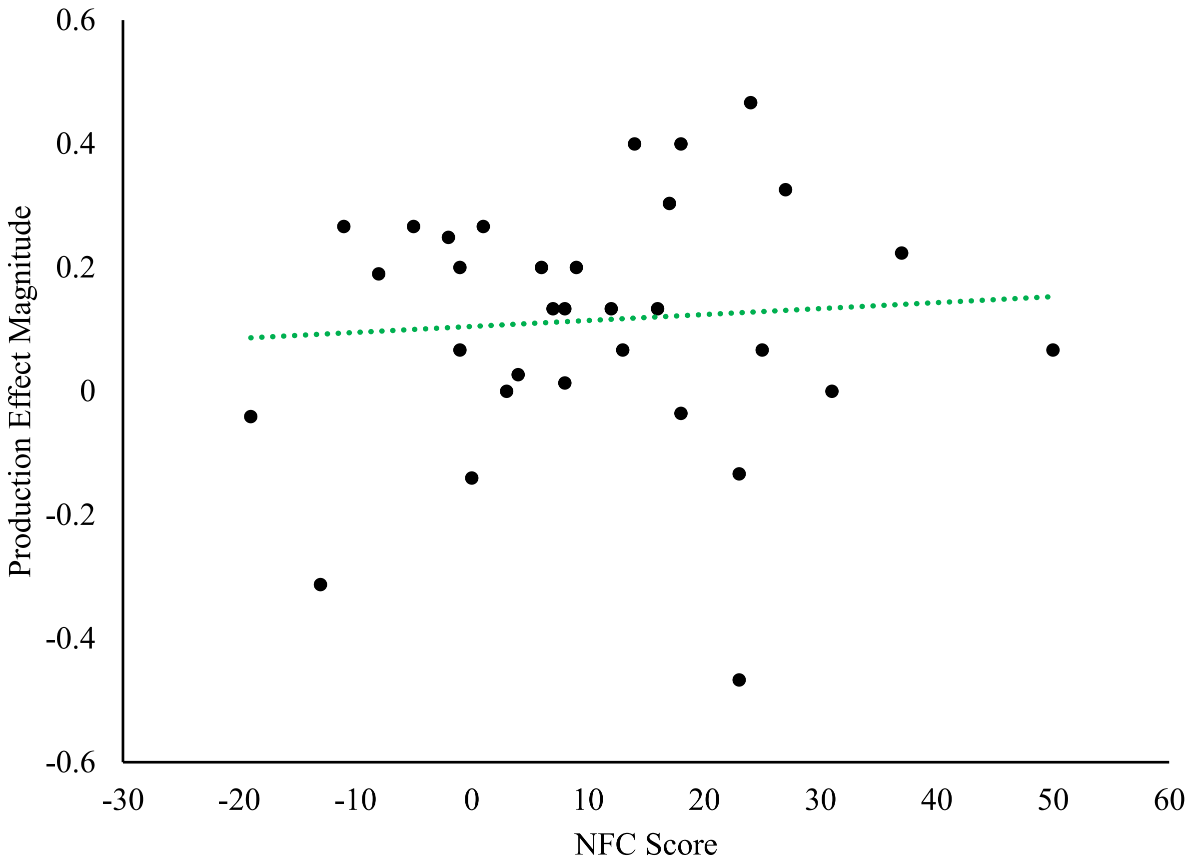
Darling test (all $p > .05$), and a visual inspection of the associated Q-Q plot all suggested that the data met the assumption of normality. As such, a Student's paired t-test was used. There was a significant difference between encoding conditions, with participants recalling significantly more aloud words ($M = 0.32$, $SD = 0.13$) than silent words ($M = 0.21$, $SD = 0.13$), t (31) = 3.24, $p <$ 0.01, $d = 0.57$.

Influence of NFC on Memory Performance

Similar to the first two experiments, the descriptive statistics regarding NFC-18 and VVIQ performance can be found in Table 1. To determine if NFC had an influence on the magnitude of the production effect, a linear regression was performed between the NFC scores and the difference scores between the encoding conditions. NFC score was not a significant predictor of the production effect's magnitude, F (1, 30) = 0.16, $p = .692$, R = .072, $R^2 = .005$, Adjusted $R^2 = -.003$ (see Figure 6). Concerning the active and baseline encoding conditions, NFC significantly predicted the mean number of aloud words that were recalled, F (1, 30) = 4.29, $p = .047$, R = .353, $R^2 = .15$, Adjusted $R^2 = .096$. However, NFC did not significantly predict the mean number of silent words that were recalled, F (1, 30) = 1.87, $p = .182$, R = .242, $R^2 = .059$, Adjusted $R^2 = .027$. But NFC score was a significant predictor of overall recall performance, F (1, 30) = 6.93, $p = .013$, R = .433, $R^2 = .019$, Adjusted $R^2 = .161$.

Figure 6

Scatterplot of the Relationship Between Production Effect Magnitude and NFC Score



As a number of these findings are nonsignificant, corresponding Bayesian analyses were also run to determine how much support could be provided to the null and alternative hypotheses. A Bayesian linear regression was performed using default priors to ascertain whether NFC substantially influenced the difference scores between the aloud and silent conditions. This analysis suggested that the null hypothesis is 2.79 times more likely than the alternative hypothesis, demonstrating inconclusive evidence in favour of NFC having no substantial effect on the magnitude of the production effect ($BF_{10} = 0.357$). Although a frequentist linear regression initially suggested a significant relationship, evidence supporting the connection between NFC and recall performance for the aloud words was also inconclusive. Namely, a

Bayesian linear regression only suggested that the alternative hypothesis was 1.64 times more likely than the null hypothesis (BF_{10} = 1.64). A Bayesian linear regression also suggested an inconclusive relationship between NFC and the average number of silent words that were recalled. Specifically, the null hypothesis appeared to be 1.47 times more likely than the alternative hypothesis (BF_{10} = 0.68). Concerning NFC and overall recall, a Bayesian linear regression with default priors suggested that the alternative hypothesis was 4.09 times more likely than the null hypothesis (BF_{10} = 4.09).

Exploratory Analyses

VVIQ score was not significantly predictive of the production effect's magnitude, F (1, 30) = 0.175, p = .679, R = .008, R^2 = .006, Adjusted R^2 = -.0273, nor was the scale a significant predictor of overall recall performance, F (1, 30) = 1.16, p = .289, R = .193, R^2 = .037, Adjusted R^2 = .005. Similarly, NFC scores were not a significant predictor of VVIQ scores, F (1, 30) = 0.01, p = .903, R = .022, R^2 = .000, Adjusted R^2 = .015. However, a corresponding Bayes inclusion factor indicated that the finding is inconclusive, with the null hypothesis being only 2.96 times more likely than the alternative hypothesis (BF_{10} = 0.34).

Similar to the first two experiments, the NFC-18 used in Experiment 3 was measured on a 9-point Likert scale. As such, the NFC data were also collapsed into a 5-point scale. Descriptive statistics involving these collapsed data can be found in Table 2.

Discussion

In Experiment 3, participants read words aloud and silently before being tested on their recall ability for all of the items. As expected, a within-subjects production effect was replicated in free recall. Specifically, participants tended to recall more words that were read aloud than words that were studied silently.

In line with established hypotheses, the magnitude of the observed production effect did not appear to have a significant relationship with NFC. However, this was not fully supported, as there was only inconclusive evidence surrounding whether NFC could predict how many words were recalled in the "aloud" or "silent" conditions, or the production effect difference score. But there did appear to be a significant relationship between participants' NFC score and their overall recall performance. Moreover, like the first two experiments, VVIQ did not significantly predict memory performance. But unlike the previous experiments, VVIQ and NFC did not appear to have any significant relationship with one another. But as this experiment lacks a full sample size, more data will need to be collected before any definitive conclusions can be drawn.

Compared to the number of drawn words and generated words that were recalled in Experiment 1 and 2, the recall performance for the aloud words in Experiment 3 was the lowest of the active encoding strategies. In addition to supporting the multimodal encoding model, the observation that fewer produced words were recalled than generated and drawn words suggests that production aids memory by facilitating item-specific processing rather than elaborative processing – with this item-specific processing leading to comparatively weaker memory performance. However, this smaller memory benefit could also be attributed to Experiment 3 having the shortest encoding interval out of the three experiments.

Chapter 5: General Discussion

The primary aim of these experiments was to determine whether NFC significantly influenced the magnitude of memory benefit imparted by various encoding strategies. The proposed outcomes of this thesis hypothesized that the drawing effect, generation effect, and production effect would be replicated. It was also expected that NFC would have an inverse relationship with the mnemonic benefits of drawing and generating. Namely, it was predicted that high-NFC individuals would benefit less from highly elaborative encoding strategies for two main reasons: 1) Individuals high in NFC would be more inclined to engage in the additional effort of spontaneous elaborative processing during baseline trials – leading to greater baseline recall. Meanwhile, individuals low in NFC were expected to simply follow task instructions and be unmotivated to expend extra cognitive effort. 2) Although high-NFC individuals may be more inclined to engage in elaborative processing during active trials as well, they would already be doing so in accordance with task instructions. A highly elaborative encoding strategy would then facilitate a comparable amount of deep processing from low-NFC participants, resulting in comparable active performance between the two NFC types. These hypotheses were tested by running drawing effect, generation effect, and production effect paradigms, and then measuring participants' NFC.

In terms of findings, the drawing effect, generation effect, and production effect were all replicated. These support the current literature surrounding the robustness of these effects in within-subjects recall (Fernandes et al., 2018; MacLeod & Bodner, 2017; McCurdy et al., 2020). In observing the mean recall performance for each of the experiments, the drawing effect paradigm elicited the highest memory performance while the production effect showed the lowest average recall scores. This trend coincides with the amount of elaborative processing theorized to be involved with each strategy, with drawing being considered highly elaborative (Fernandes et al., 2018), generation encouraging a mixture of item-specific and relational processing (Hirshman & Bjork, 1988; McCurdy et al., 2020), and production focusing more on phonetic, item-specific processing (Caplan & Guitard, in press; MacLeod et al., 2010). This would be unsurprising given that the act of elaboration during encoding allows for more complex memory traces, improving memory by providing more retrieval cues for later retention (Craik &

Lockhart, 1972; Baddeley et al., 2020). However, comparisons between the current experiments should also be made with caution, as the differences in elaborative processing are also confounded with differences in encoding interval (as well as other methodological differences, including differences in the baseline tasks). Specifically, longer encoding intervals are associated with higher memory performance in the drawing effect paradigm (Wammes et al., 2016). Additionally, self-paced generation effect paradigms have been shown to result in greater memory performance as well as larger generation effects (McCurdy et al., 2020). As such, the superior memory performance seen in the current drawing effect and generation effect experiments may also be influenced by the longer encoding intervals relative to those provided in the production effect experiment.

Although the present experiments were able to replicate the drawing effect, generation effect, and production effect, NFC level was not found to significantly predict any of these memory benefits. This contradicts the primary hypotheses for Experiment 1 and 2, but does support the prediction for Experiment 3. As such, it is possible that these strategies are effective regardless of one's NFC level. As NFC also did not significantly predict differences in baseline performances for the writing, typing, and silent conditions, this suggests that baseline memory performances do not appear to differ across NFC level in the drawing effect, generation effect, and production effect paradigms. These findings suggest three possibilities. The first is that the relationship between NFC and memory effect magnitude does exist, and the present experiments were unable to detect it due to sampling error. The second possibility is that NFC does not have a direct impact on memory effect magnitude, but rather this trait interacts with a factor that was not measured in the present studies. Finally, the third possibility is that while NFC appears to encourage elaborative processing in other scenarios (Leding, 2011; Wootan & Leding, 2015), one's NFC level may not significantly influence the effectiveness of the specific memory strategies that were explored in the present experiments.

In line with the first possibility, the present findings suggested that NFC was also not predictive of overall memory performance in the drawing effect and generation effect paradigms. While support for a relationship between NFC and overall memory performance was found in the production effect experiment, these data also consisted of a small sample size. As such, this relationship should be viewed with caution until more data are collected.

This mixed pattern between NFC and overall recall contrasts previous studies that have found greater recall performance for high-NFC individuals in comparison to low-NFC counterparts (Cacioppo et al., 1983; Kardash & Noel, 2000; Kuo et al., 2012; Leding, 2011; Peltier & Schibrowsky, 1994; Reid et al., 1995; Wootan & Leding, 2015). This difference between the current findings and past research could be partially explained by differences in study material. For instance, a number of recall studies involving NFC tend to centre around expository materials such as text passages (Cacioppo et al., 1983; Kardash & Noel, 2000) and advertising materials (Peltier & Schibrowsky, 1994; Kuo et al., 2012). As such, it is possible that these items provide more information for high-NFC individuals to deliberate over, leading to more pronounced memory differences relative to the word lists studied in the present experiments. However, the contradiction becomes more apparent when comparing the current findings with those of Leding (2011) and Wootan and Leding (2015), who also had participants study word lists.

As such, one explanation regarding the present findings concerns the potential for sampling error. Not only did the current experiments sample from a population of university students, but the sample also consisted entirely of individuals who volunteered in exchange for additional course credit. Consequently, it is likely that the current sample is unrepresentative of the full range of possible NFC scores. Although a spectrum of NFC has been demonstrated within undergraduate samples, these have also been shown to have a more limited range compared to those found in the general public (Caccippo et al., 1996). This presents the possibility that the influence of NFC on elaborative encoding strategies may still be seen in a sample with more representation in low-NFC scores. As such, it would be beneficial to run a replication of the current experiments using a sample from a more general population.

In a similar vein, participants who were lower in NFC may have been sufficiently motivated to perform well in the present experimental tasks, resulting in comparable memory performance with their high-NFC counterparts. Although motivation was not measured in the present experiments, low-NFC individuals have been shown to differ from high-NFC counterparts in terms of motivation (Amabile et al., 1994; Olsen et al., 1984). For instance, low-NFC persons are more likely to engage with cognitive problems when provided with external incentives (e.g., when given a monetary or social reward, or a reason aligning with their personal values). Meanwhile, people who score high in NFC are conceptualized as being more spontaneous, intrinsically motivated thinkers. Given that all of the present samples volunteered to participate in exchange for additional course credit, there is the potential that low-NFC participants were just as motivated to engage with the task as their high-NFC counterparts. This could have resulted in relatively equal baseline performances between high and low NFC participants, even if their reasons for being motivated may have come from different sources.

But while it is possible that the present experiments were too limited in NFC range to capture the hypothesized relationship between NFC and memory, relationships between NFC scores and elaborative processing have been observed in similar samples (Leding, 2011; Wootan

& Leding 2015). Specifically, as seen in Table 2, the current NFC scores were collapsed into a 5-point Likert scale for ease of comparison with previous NFC research. The ranges for the three experiments were 40-81, 29-78, and 45-80, respectively. Aside from the third experiment, these ranges coincide with previous research on undergraduate samples, as Leding (2011) reported a range of 34–81 while Wootan and Leding (2015) found a range of 39–84 in their sample. However, the statistical analyses used by these authors differed from those used in the present experiments. Specifically, both studies divided participants into high and low-NFC groups, then conducted ANOVAs to determine any differences. Meanwhile, the present experiments treated NFC score as a continuous variable and ran linear regressions. As such, these statistical differences could have contributed to the contrast in findings.

Additionally, although it is still possible that the hypothesized relationships may become apparent in a sample with a wider NFC range, it is also conceivable that NFC does not directly impact memory effect magnitude. Rather, there is the potential that NFC interacts with another factor that was not measured in the current experiments, such as long-term memory ability (individual differences in "the system responsible for maintaining all of the memories a person has acquired over the lifespan"; (p. 81), Unsworth, 2019). Although the relationship between NFC and long-term memory ability has yet to be directly explored, work by Unsworth and Miller (2024) suggests that the effectiveness of elaborative processing on memory may be moderated by one's long-term memory abilities. Specifically, they found that those who scored highly on measures of long-term memory ability (i.e., delayed free recall, picture source recognition, and paired associates tasks) were also more likely to benefit from deep processing in the levels of processing paradigm. As such, although NFC did not demonstrate a direct impact on the mnemonic benefits of drawing and generating, long-term memory ability may moderate this relationship. Namely, it is possible that individuals who are both high in NFC and long-term memory ability would have both the motivation and the capability to remember more baseline items. As such, they might perform differently in the drawing effect and generation effect paradigms when compared to those who are low in both of these traits.

Finally, it is possible that one's NFC level simply does not pose a substantial influence over the benefits of drawing and generating. A similar finding in the levels of processing paradigm was reported by Wootan and Leding (2015), who found that while high-NFC participants recalled more target words than low-NFC counterparts in both the deeply and shallowly processed conditions, both NFC groups showed comparable performance differences between the deep and shallow item conditions. In other words, the magnitude of the levels of processing effect was similar between the two NFC groups even though the high-NFC participants demonstrated greater overall recall. With regard to the present experiments, it is feasible that the drawing and generation effect paradigms are just too cognitively demanding to allow the high-NFC participants to engage in extra elaborative processing.

Finally, a curious finding was the relationship demonstrated between VVIQ and NFC. In Experiment 1, high NFC scores were predictive of high VVIQ scores. However, in Experiment 2, the inverse was found – high NFC scores predicted low VVIQ scores. The opposing directions of these findings are surprising as both NFC and VVIQ scores are considered to be measures of trait-level differences. Given the contradiction between the two findings, they do not easily fit into what has been previously found surrounding NFC and visual imagery. Although the NFC and VVIQ have yet to be directly compared to other studies, NFC does not appear to have a significant relationship with visual processing preferences, instead demonstrating a preference for verbal information (Sojka & Giese, 2001; Venkatraman et al., 1990). However, Huang and

Ha (2020) found that NFC moderates the relationship between advertising fluency and mental imagery about the advertisement's brand. They found that when brand advertisements were perceived as fluent, individuals high in NFC were more likely to score higher on the Consumption Vision Scale, a measure of mental imagery (Walters et al., 2007). However, the Consumption Vision Scale is rather different from the VVIQ, containing items such as "It was easy for me to imagine wearing this brand product", while the VVIQ asks participants to imagine different landscapes and everyday scenes.

As such, it is possible that the current findings surrounding NFC and VVIQ may be due to alternative factors, such as social desirability biases and priming. NFC has a debated relationship with social desirability, with Cacioppo and Petty's (1982) original study finding both a lack of a significant relationship in their third experiment, as well as a weak positive relationship in their fourth study. Other studies have replicated both the lack (Fletcher et al., 1986; Petty & Jarvis, 1996) and the presence of this correlation (Olson et al., 1984). However, if it is the case that individuals higher in NFC are more susceptible to social desirability, it is possible that the high-NFC individuals felt obligated to respond positively to the VVIQ after participating in a study about drawing. Individuals high in NFC are also thought to be more prone to priming (Petty & Jarvis, 1996). As such, high-NFC participants in Experiment 1 may have been more susceptible to unintentional priming. Specifically, by presenting the VVIQ after the drawing effect paradigm, high-NFC participants may have been primed by the visual nature of the memory task, resulting in higher VVIQ scores as NFC scores increased. As Experiment 2 dealt with generating rather than drawing, it is conceivable that the high-NFC participants answered with less bias, leading to the inverse relationship that was observed.

Limitations

In addition to the limitations in NFC range, another limitation concerns the online format of Experiment 2. Although the implementation of a "type" condition helped to check for instruction compliance, this experiment was still conducted unsupervised. Despite recruiting a sample size that was comparable to other NFC studies, the Bayes factors still suggest inconclusive data, supporting the notion that the online experiment contained extra noise. As such, the current findings should be interpreted with caution and more data should be collected before drawing any substantial inferences. Another limitation revolves around the limited sample size collected for Experiment 3. While the first two experiments sampled over 100 participants each, the third experiment only recruited 41 students due to time constraints. Although the current sample size is representative of those found in within-subjects production effect studies (MacLeod et al., 2010), studies exploring NFC often require much more to detect any substantial relationships. As such, Experiment 3 is underpowered – a fact that is seen in both the previously mentioned NFC range (45-80), as well as the inconclusive Bayesian analyses conducted for this study. Therefore, caution must be taken when interpreting the relationships involving NFC for this experiment and more data must be collected before any definitive inferences can be made.

Future Directions

Future studies could explore the relationship between NFC and active encoding strategies in a more general sample in order to get a more diverse range of NFC scores. This would address the restricted range and uniform participant samples tested in the current series of experiments. Additionally, NFC and its potential role in elaborative processing could be explored with either consecutive testing or increased retention intervals. Specifically, both Leding (2011) and Wootan and Leding (2015) found that with multiple memory tests, high-NFC individuals would show increased true and false recall relative to their low-NFC counterparts. Leding and colleagues proposed that this was due to the high-NFC participants having stronger semantic memory traces following the eventual decay of more item-specific detail memories. As such, although no substantial patterns were observed between NFC and the memory paradigms explored in the current experiments, a relationship could be investigated with manipulations of retention intervals. Any NFC-related differences in memory performance may be more salient with longer retention intervals, as low-NFC participants may be more likely to experience quicker memory trace decay relative to their high-NFC counterparts.

Conclusion

The current thesis sought to explore whether NFC impacts the mnemonic benefits of encoding strategies. Although the drawing, generation, and production effects were replicated, the proposed relationship between NFC and the relative magnitude of these effects was not clearly observed, with a number of inconclusive statistical findings. These indeterminate results could be attributed to both sampling error and insufficient sample size. Specifically, given the recruitment method employed in each experiment, these findings may be skewed toward high-NFC participants, limiting the range of the data. At the same time, the present studies would benefit from more data in order to draw more definitive conclusions. However, if it is true that NFC has no bearing on the benefits of drawing, generation, and production, then these encoding strategies can be employed without having to consider individual differences in NFC level.

References

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: Assessing Intrinsic and Extrinsic Motivational Orientations. *Journal of Personality and Social Psychology*, *66*(5), 950–967. https://doi.org/10.1037/0022-3514.66.5.950

Baddeley, A. D., Eysenck, M. W., & Anderson, M. (2020). *Memory* (3rd Edition). Psychology Press.

Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology.*, *15*(5), 977–989. https://doi.org/10.1037/0278-7393.15.5.977

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Berzonsky, M. D., & Sullivan, C. (1992). Social-Cognitive Aspects of Identity Style: Need for Cognition, Experiential Openness, and Introspection. Journal of Adolescent Research, 7(2), 140-155. https://doi.org/10.1177/074355489272002

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review.*, *21*(1), 149–154. https://doi.org/10.3758/s13423-013-0485-1

Booth-Butterfield, M., & Booth-Butterfield ,S. (1990). Conceptualizing Affect as Information in Communication Production. *Human Communication Research*, *16*(4), 451–476. https://doi.org/10.1111/j.1468-2958.1990.tb00219.x

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment.*, *48*(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin.*, *119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology.*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Clyde, M. A. (2017). *BAS: Bayesian Adaptive Sampling for Bayesian Model Averaging*. [R package]. Retrieved from https://cran.r-project.org/package=BAS.

Clyde, M. A., Ghosh, J., & Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics, 20*, 80-101.

Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *Journal of Abnormal and Social Psychology*, *51*(2), 291-294

Conway, M. A., & Gathercole, S. E. (1990). Writing and long-term memory: Evidence for a “translation” hypothesis. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *42*(3), 513–527. https://doi.org/10.1080/14640749008401235

Conway, M. A., & Gathercole, S. E. (1990). Writing and long-term memory: Evidence for a “translation” hypothesis. The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 42(3), 513–527. https://doi.org/10.1080/14640749008401235

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior.*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology.*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Eigenberger, M. E., Critchley, C., & Sealander, K. A. (2007). Individual differences in epistemic style: A dual-process perspective. *Journal of Research in Personality*, *41*(1), 3–24. https://doi.org/10.1016/j.jrp.2006.01.003

Elphinstone, B., Farrugia, M., Critchley, C., & Eigenberger, M. E. (2014). Examining the measurement of epistemic style: The development and validation of the Epistemic Preference Indicator-Revised. *Personality and Individual Differences*, *58*, 101–105. https://doi.org/10.1016/j.paid.2013.10.013

Evans, J., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Fawcett, J. M., Baldwin, M. M., Whitridge, J. W., Swab, M., Malayang, K., Hiscock, B., Drakes, D. H., & Willoughby, H. V. (2023). Production Improves Recognition and Reduces Intrusions in Between-Subject Designs: An Updated Meta-Analysis. *Canadian Journal of Experimental Psychology*, *77*(1), 35–44. https://doi.org/10.1037/cep0000302

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory (Hove)*, *20*(7), 655–666. https://doi.org/10.1080/09658211.2012.693510

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The Surprisingly Powerful Influence of Drawing on Memory. *Current Directions in Psychological Science : A Journal of the American Psychological Society*, *27*(5), 302–308. https://doi.org/10.1177/0963721418755385

Fletcher, G. J. O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual differences measure. *Journal of Personality and Social Psychology.*, *51*(4), 875–884. https://doi.org/10.1037/0022-3514.51.4.875

Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (2014). Production improves memory equivalently following elaborative vs non-elaborative processing. *Memory (Hove)*, *22*(5), 470–480. https://doi.org/10.1080/09658211.2013.798417

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*(7), 1046–1055. https://doi.org/10.3758/s13421-012-0210-8

Graham, L. M. (2007). Need for cognition and false memory in the Deese–Roediger–McDermott paradigm. *Personality and Individual Differences*, *42*(3), 409–418. https://doi.org/10.1016/j.paid.2006.07.012

Heppner, P. P., Reeder, B. L., & Larson, L. M. (1983). Cognitive variables associated with personal problem-solving appraisal: Implications for counseling. *Voices of the Forgotten Half: The Role of Social Class in the School-to-Work Transition*, *30*(4), 537–545. https://doi.org/10.1037/0022-0167.30.4.537

Hill, B. D., Foster, J. D., Elliott, E. M., Shelton, J. T., McCain, J., & Gouvier, Wm. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. *Journal of Research in Personality*, *47*(1), 22–25. https://doi.org/10.1016/j.jrp.2012.11.001

Hirshman, E., & Bjork, R. A. (1988). The Generation Effect: Support for a Two-Factor Theory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(3), 484–494. https://doi.org/10.1037/0278-7393.14.3.484

Hourihan, K. L., & Churchill, L. A. (2020). Production of Picture Names Improves Picture Recognition. *Canadian Journal of Experimental Psychology*, *74*(1), 35–43. https://doi.org/10.1037/cep0000185

Hourihan, K. L., & MacLeod, C. M. (2007). Capturing conceptual implicit memory: The time it takes to produce an association. *Memory & Cognition*, *35*(6), 1187–1196. https://doi.org/10.3758/BF03193592

Hunt, R. R. (2013). Precision in Memory Through Distinctive Processing. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, *22*(1), 10–15. https://doi.org/10.1177/0963721412463228

Hunt, R., & McDaniel, M. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*(4), 421–445. https://doi.org/10.1006/jmla.1993.1023

Hussey, I., & Hughes, S. (2020). Hidden Invalidity Among 15 Commonly Used Measures in Social and Personality Psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 166–184. https://doi.org/10.1177/2515245919882903

Huang, R., & Ha, S. (2021). The role of need for cognition in consumers' mental imagery: a study of retail brand's Instagram. *International Journal of Retail & Distribution Management.*, *49*(2), 242–262. https://doi.org/10.1108/IJRDM-04-2020-0146

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82*(3), 472–481. https://doi.org/10.1037/h0028372

Jamieson, R. K., & Spear, J. (2014). The Offline Production Effect. *Canadian Journal of Experimental Psychology*, *68*(1), 20–28. https://doi.org/10.1037/cep0000009

JASP Team (2018). *JASP*. [Computer software]. Retrieved from https://jasp-stats.org.

Jonker, T. R., Wammes, J. D., & MacLeod, C. M. (2019). Drawing Enhances Item Information but Undermines Sequence Information in Memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*(4), 689–699. https://doi.org/10.1037/xlm0000610

Kardash, C. M., & Noel, L. K. (2000). How Organizational Signals, Need for Cognition, and Verbal Ability Affect Text Recall and Recognition. *Contemporary Educational Psychology*, *25*(3), 317–331. https://doi.org/10.1006/ceps.1999.1011

Kelley, M. R., Chapman-Orr, E. K., Calkins, S., & Lemke, R. J. (2019). Generation and Retrieval Practice Effects in the Classroom Using PeerWise. *Teaching of Psychology.*, *46*(2), 121–126. https://doi.org/10.1177/0098628319834174

Kuo, J.-C., Horng, D.-J., Lin, C.-L., & Lee, S.-H. (2012). The Causal Relationship Between Need for Cognition and Advertising Recall. *Social Behavior and Personality*, *40*(6), 1025–1032. https://doi.org/10.2224/sbp.2012.40.6.1025

Leding, J. K. (2011). Need for Cognition and false recall. *Personality and Individual Differences*, *51*(1), 68–72. https://doi.org/10.1016/j.paid.2011.03.017

Love, J., & Moreno, V. (2024, May 28). *jamovi library*. Jamovi. https://www.jamovi.org/library.html

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The Production Effect: Delineation of a Phenomenon. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *The British Journal of Psychology*, *64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review*, *27*(6), 1139–1165. https://doi.org/10.3758/s13423-020-01762-3

McElroy, L. A., & Slamecka, N. J. (1982). Memorial Consequences of Generating Nonwords: Implications for Semantic-Memory Interpretations of the Generation Effect. *Journal of Verbal Learning and Verbal Behavior*, *21*(3), 249-. https://doi.org/10.1016/S0022-5371(82)90593-X

Meade, M. E., Wammes, J. D., & Fernandes, M. A. (2018). Drawing as an Encoding Tool: Memorial Benefits in Younger and Older Adults. *Experimental Aging Research*, *44*(5), 369–396. https://doi.org/10.1080/0361073X.2018.1521432

Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not extrinsic context. *Quarterly Journal of Experimental Psychology (2006)*, *64*(8), 1543–1562. https://doi.org/10.1080/17470218.2011.562980

Olson, K., Camp, C., & Fuller, D. (1984). Curiosity and Need for Cognition. *Psychological Reports.*, *54*(1), 71–74. https://doi.org/10.2466/pr0.1984.54.1.71

Peltier, J., & Schibrowsky, J. (1994). Need for cognition, advertisement viewing time and memory for advertising stimuli. *Advances in Consumer Research*, *21*, 244–250.

Petty, R. E., Briñol, P., Loersch, C., & McCaslin, M. J. (2009). Need for cognition. In M. R. Leary & R. H. Hoyle, *Handbook of individual differences in social behaviour* (pp. 318-329). The Guilford Press.

Petty, R. E., & Jarvis, W. B. G. (1996). An individual differences perspective on assessing cognitive processes. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 221–257). Jossey-Bass/Wiley.

Pinard, F., Vanneste, S., Bouazzaoui, B., Fay, S., Isingrini, M., Angel, L., Borella, E., & Taconnat, L. (2023). Personality and organizational strategies in episodic memory: The role of Openness to experience on recall and subjective organization in young and older adults. *L'Année Psychologique.*, *123*(3), 513–537. https://doi.org/10.3917/anpsy1.233.0513

Putnam, A. L., Ozubko, J. D., MacLeod, C. M., & Roediger, H. L. (2014). The production effect in paired-associate learning: Benefits for item and associative information. *Memory & Cognition*, *42*(3), 409–420. https://doi.org/10.3758/s13421-013-0374-x

Quinlan, C. K., & Taylor, T. L. (2019). Mechanisms underlying the production effect for singing. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale.*, *73*(4), 254–264. https://doi.org/10.1037/cep0000179

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reid, J. C., Klachko, D. M., Kardash, C. A. M., Robinson, R. D., Scholes, R., & Howard, D. (1995). Why people don't learn from diabetes literature: influence of text and reader characteristics. *Patient Education and Counseling*, *25*(1), 31–38. https://doi.org/10.1016/0738-3991(94)00688-I

Reisberg, D. (2018). *Cognition : Exploring the science of the mind* (7th edition.). W. W. Norton & Company.

Roberts, B. R. T., & Wammes, J. D. (2021). Drawing and memory: Using visual production to alleviate concreteness effects. *Psychonomic Bulletin & Review*, *28*(1), 259–267. https://doi.org/10.3758/s13423-020-01804-w

Roediger, H. L., & McDermott, K. B. (1995). Creating False Memories: Remembering Words Not Presented in Lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Sadowski, C. J. & Gulgoz, S. (1992). Internal consistency and test-retest reliability of the need for cognition scale. *Perceptual and Motor Skills.*, *74*(2), 610–610. https://doi.org/10.2466/PMS.74.2.610-610

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Sojka, J. Z., & Giese, J. L. (2001). The influence of personality traits on the processing of visual and verbal information. *Marketing Letters.*, *12*(1), 91–106. https://doi.org/10.1023/A:1008132422468

Soraci, S. A., Carlin, M. T., Toglia, M. P., Chechile, R. A., & Neuschatz, J. S. (2003). Generative processing and false memories: When there is no cost. *Journal of Experimental Psychology*, *29*(4), 511–523. https://doi.org/10.1037/0278-7393.29.4.511

Soubelet, A., & Salthouse, T. A. (2010). The role of activity engagement in the relations between Openness/Intellect and cognition. *Personality and Individual Differences.*, *49*(8), 896–901. https://doi.org/10.1016/j.paid.2010.07.026

Taconnat, L., Pinard, F., Vanneste, S., Bouazzaoui, B., Fay, S., Martinez, L., Alibran, E., & Geraci, L. (2022). Personality traits affect older adults' memory differently depending on the environmental support provided at encoding. *Personality and Individual Differences.*, *191*. https://doi.org/10.1016/j.paid.2022.111572

The jamovi project (2023). jamovi (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org

Thiede, K. W., Wright, K. L., Hagenah, S., Wenner, J., Abbott, J., & Arechiga, A. (2022). Drawing to improve metacomprehension accuracy. *Learning and Instruction*, *77*, 101541-. https://doi.org/10.1016/j.learninstruc.2021.101541

Thompson, M. M., & Zanna, M. P. (1995). The Conflicted Individual: Personality-Based and Domain Specific Antecedents of Ambivalent Social Attitudes. *Journal of Personality.*, *63*(2), 259–288. https://doi.org/10.1111/j.1467-6494.1995.tb00810.x

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall Accuracy and Illusory Memories: When More is Less. *Memory, 7*(2), 233–256. https://doi.org/10.1080/741944069

Tran, S. H. & Fernandes., M. A.(2023) Age differences in effectiveness of encoding techniques on memory. *Aging, Neuropsychology, and Cognition*, 1-17.

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin.*, *145*(1), 79–139. https://doi.org/10.1037/bul0000176

Unsworth, N., & Miller, A. L. (2024). An examination of individual differences in levels of processing. *Memory.*, 1–12. https://doi.org/10.1080/09658211.2024.2355309

Wammes, J. D., Jonker, T. R., & Fernandes, M. A. (2019). Drawing improves memory: The importance of multimodal encoding context. *Cognition,* 191. https://doi.org/10.1016/j.cognition.2019.04.024

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *The Quarterly Journal of Experimental Psychology, 69*(9), 1752–1776. https://doi.org/10.1080/17470218.2015.1094494

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2017). Learning terms and definitions: Drawing and the role of elaborative encoding. *Acta Psychologica*, *179*, 104–113. https://doi.org/10.1016/j.actpsy.2017.07.008

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2018). Creating a Recollection-Based Memory Through Drawing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(5), 734–751. https://doi.org/10.1037/xlm0000445

Walters, G., Sparks, B., & Herington, C. (2007). The Effectiveness of Print Advertising Stimuli in Evoking Elaborate Consumption Visions for Potential Travelers. *Journal of Travel Research*, *46*(1), 24–34. https://doi.org/10.1177/0047287507302376

Wootan, S. S., & Leding, J. K. (2015). Need for Cognition and false memory: Can one's natural processing style be manipulated by external factors? *The American Journal of Psychology, 128*(4), 459–468. https://doi.org/10.5406/amerjpsyc.128.4.0459Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." Journal of Open Source Software, *4*(43), 1686. doi:10.21105/joss.01686

Zerna, J., Strobel, A., & Strobel, A. (2024). The Role of Need for Cognition in Well-Being – Review and Meta-Analyses of Associations and Potentially Underlying Mechanisms. *Collabra. Psychology*, *10*(1), 20–38. https://doi.org/10.1525/collabra.92885

Appendix A

Ethics Approval Letter

MEMORIAL UNIVERSITY

Interdisciplinary Committee on
Ethics in Human Research (ICEHR)

St. John's, NL Canada A1C 5S7
Tel: 709 864-2561 icehr@mun.ca
www.mun.ca/research/ethics/humans/icehr

ICEHR Number:	**20240642-SC**
Approval Period:	October 3, 2023 – October 31, 2024
Funding Source:	
Responsible Faculty:	Dr. Kathleen Hourihan Department of Psychology
Title of Project:	*Personality factors in memory strategies*

October 3, 2023

Ms. Justine Yick
Department of Psychology, Faculty of Science
Memorial University

Dear Ms. Yick:

Thank you for your correspondence addressing the issues raised by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) for the above-named research project. ICEHR has re-examined the proposal with the clarifications and revisions submitted, and is satisfied that the concerns raised by the Committee have been adequately addressed. In accordance with the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2)*, the project has been granted *full ethics clearance* for **one year**. ICEHR approval applies to the ethical acceptability of the research, as per Article 6.3 of the *TCPS2*. Researchers are responsible for adherence to any other relevant University policies and/or funded or non-funded agreements that may be associated with the project. If funding is obtained subsequent to ethics approval, you must submit a Funding and/or Partner Change Request to ICEHR so that this ethics clearance can be linked to your award.

The *TCPS2* **requires** that you **strictly adhere to the protocol and documents as last reviewed** by ICEHR. If you need to make additions and/or modifications, you must submit an Amendment Request with a description of these changes, for the Committee's review of potential ethical concerns, before they may be implemented. Submit a Personnel Change Form to add or remove project team members and/or research staff. Also, to inform ICEHR of any unanticipated occurrences, an Adverse Event Report must be submitted with an indication of how the unexpected event may affect the continuation of the project.

The *TCPS2* **requires** that you submit an Annual Update to ICEHR before **October 31, 2024**. If you plan to continue the project, you need to request renewal of your ethics clearance and include a brief summary on the progress of your research. When the project no longer involves contact with human participants, is completed and/or terminated, you are required to provide an annual update with a brief final summary and your file will be closed. All post-approval ICEHR event forms noted above must be submitted by selecting the ***Applications: Post-Review*** link on your Researcher Portal homepage. We wish you success with your research.

Yours sincerely,

[redacted]

Alyson Byrne, Ph.D.
Vice-Chair, Interdisciplinary Committee on
Ethics in Human Research

AB/bc

cc: Supervisor – Dr. Kathleen Hourihan, Department of Psychology

Note: Any signatures have been redacted for privacy.

Appendix B

Stimuli List, Demographic Questions, and Measures

Table B1

Stimuli List Used in all Three Experiments

Target	Definition
airplane	a winged flying vehicle that people use to travel great distances
couch	a long, cushioned piece of furniture used for seating multiple people
kite	a toy consisting of a light frame with thin material stretched over it, flown in the wind at the end of a long string
ruler	a rectangular tool used to measure lengths and draw straight lines
ant	a type of insect that has no wings, lives in colonies, and often collects crumbs
cow	a farm animal that produces dairy milk
knife	a sharp utensil used for chopping and slicing food
sailboat	a water vehicle with fabric attached to a mast that catches the wind
axe	a tool used for chopping wood
desk	a type of table commonly found in classrooms and offices
ladder	a structure, made of a series of rungs, used for climbing
scissors	a tool with two blades is often used for cutting paper
balloon	an air-filled decoration commonly found at birthday parties
doll	a children's toy that often resembles a baby or a girl

lamp	an electrical device that has a lightbulb and a shade
screwdriver	a hand tool, with either a flat or cross-shaped head, used for rotating fasteners into wood
banana	a long yellow fruit that has a peel
door	an entrance to a room or building which typically has hinges and a handle/knob
lemon	a round yellow fruit that tastes sour
sheep	a farm animal that produces wool
bee	a winged insect that makes honey
drum	a percussion instrument that makes sounds when hit with either hands or sticks
lion	a large wild cat that is known for the mane around its head
shoe	a common type of footwear worn outside
beetle	a round insect with wings and a hard shell (also a type of Volkswagen)
duck	a common aquatic bird that quacks
lips	the parts of the face that make up the opening of the mouth
skirt	an open garment worn on the lower half and is typically associated with more feminine styles
blouse	a term for a dressy women's top
ear	body part used for hearing
monkey	a type of primate with a long tail that lives in trees

spider	an arthopod with eight legs that spins webs to catch prey
boot	a taller type of footwear used to protect from inclement weather
elephant	a large grey animal with big ears, tusks, and a trunk
mushroom	a type of fungi that has a domed cap and a stalk
spoon	a type of round utensil used for eating, scooping, and stirring
broom	a cleaning tool used for sweeping
flute	a high-pitched wind instrument that is made of metal and has holes and keys to adjust notes
owl	a nocturnal bird of prey with large eyes and a small hooked beak
stool	a small seat that doesn't have a back or armrests
butterfly	an insect that feeds on nectar and has large, brightly coloured wings
fork	a type of utensil with prongs used to lift and hold food when eating
pants	a type of clothing that covers the waist and each leg
stove	a kitchen appliance used for cooking
camel	a four-legged desert animal with a hump on its back
frog	a green amphibian that hops around and eats flies
peanut	a common salted snack that is also a common allergen
strawberry	a small red fruit that is covered in small seeds, commonly used to make jam
cannon	a large piece of artillery that fires a large metal ball

giraffe	a large four-legged herbivore with a long neck and spots
pear	a sweet, yellowish-green fruit that has a wider base and a narrower top
sweater	a thicker long sleeve top that is often knitted
carrot	a long orange root vegetable
glove	a hand-covering garment, designed to protect and provide warmth to the fingers and hand
penguin	a flightless, aquatic bird that is black and white
toaster	a kitchen appliance that makes slices of bread warm and crispy
cat	a small, four-legged pet known for its meows, purrs, soft fur, and whiskers
grapes	small round purple fruit that grow in clusters, commonly used to make wine
pepper	Often found in a shaker on dining tables next to salt
trumpet	a brass musical instrument with a flared bell and three buttons, played by blowing air through a mouthpiece
caterpillar	the larval stage of a butterfly known for crawling around and eating vegetation
guitar	a common six-stringed instrument that is strummed
pig	a farm animal with a rounded body, a snout, and a curly tail
turtle	a slow-moving reptile that can hide in its shell
cherry	a small red fruit that has a long green stem and a pit inside
hammer	a hand tool commonly used for driving nails into surfaces

pineapple	a tropical fruit with spiky, green and brown skin and spiked leaves on top
violin	a stringed musical instrument played with a bow used to produce notes
clock	a timekeeping device that has hands and a circular face
harp	a stringed musical instrument played by angels
pumpkin	a round, orange gourd carved on Halloween
wagon	a four-wheeled cart that's pulled with a handle
coat	a heavier garment worn to keep warm during cold weather, often extending below the waist
jacket	a lighter piece of outerwear, designed for warmth
rabbit	a small, furry, long-eared mammal known for hopping
whistle	tool used by sports referees
corn	a type of food that has many yellow kernels on a cob
kettle	a container or device used for boiling water
rooster	a male chicken that typically calls when the sun rises
wrench	a hand tool used for turning nuts and bolts

The full stimuli list consisted of 80 concrete nouns. For each participant, 30 items from this list were randomly selected for study. The definitions were only used in Experiment 2.

Table B2

Demographic Questions used in all Three Experiments

1. What is your gender?	6. Have you ever taken any formal drawing classes outside of a high school curriculum?
2. What is your age in years?	7. Do you consider drawing one of your hobbies?
3. Which hand do you normally write with?	8. Do you consider yourself an artist?
4. How fluent is your writing ability in English?	9. Are there any other comments you would like to share about your drawing skills/experience?
5. How fluent is your speaking ability in English?	10. Were there any issues that would require us to reconsider using your experimental data (e.g., mixed up the instructions, sleepiness, noisy environment, computer lag, etc.)? If so, please describe them below. There are no consequences to experiencing issues and reporting them helps us better understand our results!

Table B3

The NFC-18 Scale

1. I would prefer complex to simple problems	7. I only think as hard as I have to.	13. I prefer my life to be full of puzzles I must solve.
2. I like to have the responsibility of handling a situation that requires a lot of thinking.	8. I prefer to think about small, daily projects to long-term ones.	14. The notion of thinking abstractly appeals to me.
3. Thinking is not my idea of fun.	9. I like tasks that require little thought once I've learned them	15. I would prefer a task that is intellectual, difficult, and important to one that somewhat important but does not require much thought.
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.	10. The idea of relying on thought to make my way to the top appeals to me.	16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.
5. I try to anticipate and avoid situations where there is a likely chance I will have to think in depth about something.	11. I really enjoy a task that involves coming up with new solutions to problems.	17. It's enough for me that something gets the job done; I don't care how or why it works.
6. I find satisfaction in deliberating hard and for long hours.	12. Learning new ways to think doesn't excite me very much.	18. I usually end up deliberating about issues even when they do not affect me personally.

Participants in this experiment rated their agreement with each statement using 9-point Likert scales (+4 = "Very strongly agree", -4 = "Very strongly disagree").

Table B4

The Vividness and Visual Imagery Questionnaire (VVIQ)

For items 1-4, think of a relative or friend whom you frequently see (but who is not with you at present) and consider carefully the picture that comes before your mind's eye.	**For items 5-8, visualize a rising sun. Consider carefully the picture that comes before your mind's eye.**	**For items 9-12, think of the front of a shop that you often go to. Consider the picture that comes before your mind's eye.**	**Finally, think of a country scene which involves trees, mountains, and a lake. Consider the picture that comes before your mind's eye.**
1. The exact contour of face, head, shoulders, and body.	5. The sun is rising above the horizon into a hazy sky.	9. The overall appearance of the shop from the opposite side of the road.	13. The contours of the landscape.
2. Characteristic poses of head, attitudes of body, etc.	6. The sky clears and surrounds the sun with blueness.	10. A window display including colours, shapes, and details of individual items for sale.	14. The colour and shape of the trees.
3. The precise carriage, length of step, etc. in walking.	7. Clouds. A storm blows up, with flashes of lightning.	11. You are near the entrance. The colour, shape, and details of the door.	15. The colour and shape of the lake.
4. The different colours worn in some familiar clothes.	8. A rainbow appears.	12. You enter the shop and go to the counter. The counter assistant serves you. Money changes hands.	16. A strong wind blows on the trees and on the lake causing waves.

Participants in these experiments rated how vividly they could imagine each statement using 5-point Likert scales (1 = "Perfectly clear and as vivid as normal vision"; 5 = "No image at all, you only 'know' that you are thinking of the object").