ENHANCING THE PRODUCTION EFFECT IN MEMORY: SINGING, UNDERLYING MECHANISMS, AND APPLICATIONS

by

Chelsea Karmel Quinlan

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
December 2017

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The production effect is the finding that memory performance is better for words that are produced (i.e., read aloud) compared to words that are not produced (i.e., read silently). This dissertation aimed to expand previous research by investigating: 1) whether alternate forms of vocal production have a greater impact on the production effect than reading aloud in a normal voice and 2) the possible mechanisms underlying any influence of alternate forms of vocal production on the production effect. In Experiments 1-3, we found evidence of a graded pattern of memory performance: Both reading items aloud loudly and singing items at study resulted in greater subsequent memory performance than did reading items aloud in a normal voice, with singing items at study resulting in even greater memory performance than reading aloud loudly. In Experiments 4 through 6, we examined possible mechanisms underlying the greater production effect for singing versus reading aloud. Our results provided evidence against three potential explanations for the greater production effect for singing versus reading aloud including a bizarreness explanation, differences in production duration, and differences in trace memory strength. Taken together, the findings from this dissertation provide evidence that alternative forms of vocal production, such as singing and reading aloud loudly, have a greater impact on memory performance than reading aloud in a normal voice. Our findings also provide strong support for a distinctiveness account of the production effect, emphasizing that the number and type of potential distinct elements available at encoding is likely associated with subsequent memory performance at test and consequently with the magnitude of the production effect (i.e., the greater the number and type of distinct elements available at encoding, the greater the magnitude of the production effect).

LIST OF ABBREVIATIONS USED

FGS	Faculty of Graduate Studies
Dal	Dalhousie University
ANOVA	Analysis of Variance
PET	Positron Emission Tomography
EEG	Electroencephalography
AD	Alzheimer's Disease
DSM-5	Diagnostic and Statistical Manual of Mental Disorders - 5th Edition

ACKNOWLEDGEMENTS

Foremost, I would like to thank my supervisor, Dr. Tracy Taylor-Helmick, for her continuous belief in me and her endless support and guidance during the production of my dissertation, as well as during the many research projects preceding this dissertation. We met over 10 years ago, and it was at that time that she ignited my passion for research in psychology. She has taught me so much over the years and I will forever admire her as an advisor, researcher, and friend.

I would like to thank my committee members, Dr. Raymond Klein and Dr. Shannon Johnson for their expertise, guidance, and insightful feedback on my dissertation. I am likewise thankful for the support, guidance, and friendship of Jonathan Fawcett, who has been there for me every step of the way throughout this journey. I would also like to thank Dr. Colin MacLeod for introducing me to the production effect and for his thoughtful feedback on the manuscript included in this dissertation.

In addition, I would like to acknowledge Carl Helmick for his assistance with software programming, as well as Kathy Otten and Ian Palmer for their help with data collection. I also would like to acknowledge Dr. Tracey Scott and Dr. Krista Barney for recognizing my potential and gently pushing me to finish my doctoral degree. I could not have done this without their incredible belief in me.

Finally, I cannot thank my husband, Daniel, and son, Finley, enough for their unwavering support and love throughout the ups and downs that accompany graduate school. In addition, I would be remiss without mentioning my family and friends, who have also been there to support and guide me along the way.

CHAPTER 1: INTRODUCTION

The following chapter is based in part upon the manuscript entitled *"The production effect: A meta-analytic review"*, which was previously submitted for publication to Memory in September 2013. Although Dr. Jonathan M. Fawcett was a co-author for this manuscript, Chelsea K. Quinlan was the primary contributing author to this manuscript; she compiled and interpreted the data for this manuscript and also produced the first initial draft as well as all major revisions prior to submission.

1.1 INTRODUCTION

Memory has been conceptualized in a number of ways. For instance, Oscar Wilde (1895, pp. 25) defined memory as “the diary that we all carry about with us.” Sternberg (1999) defines memory as “the means by which we draw on our past experiences in order to use this information in the present”, whereas Matlin (2005) states that “memory is the process of maintaining information over time.” Regardless of the specific definition of memory, it tends to be conceptualized as a constructive process that involves one or more of the following: the encoding, storage, and/or retrieval of information and stimuli. Encoding is the process whereby visual, acoustic, and semantic information in the environment is received and processed; storage involves creating a durable memory representation of the encoded information; and retrieval is the process of locating stored information and bringing it into awareness so that it can be used.

Many different variables or manipulations can influence the manner in which information and stimuli are encoded and stored in memory as well as retrieved from memory. For example, generating (e.g., stem completion) or enacting (e.g., performing the action) items at study tends to produce greater memory performance than not generating (i.e., the generation effect; Slamecka & Graf, 1978; also, see Bertsch, Pesta, Wiscott, & McDaniel, 2007, for a review) or enacting items (i.e., the enactment effect; Cohen, 1981; Engelkamp & Krumnacker, 1980). In addition, studying pictures produces greater memory performance than studying words (i.e., the picture superiority effect; Paivio, 1971). Similarly, items that are processed on a more deep level (e.g., semantic processing) at study are remembered better than items that are processed on a more shallow level (e.g., phonemic or orthographic processing; Craik & Lockhart, 1972). The von Restorff effect is the finding that an item that is different or isolated from the other items on the study list is better remembered (see Hunt, 1995).

In addition to manipulations at study, which can affect the encoding and later retrieval of information, there are several manipulations at test that can affect the retrieval of encoded information. These include but are not limited to direct versus indirect tests, cued versus uncued recall, and recall versus recognition. Furthermore, it is important to consider the interaction of the encoding and retrieval of information. For example, memory tends to be greater when there is a “match” between the environment at study and test (i.e., the processes used at study are appropriate for those required at test) compared to when there is a “mismatch” between the environment at study and test (i.e., context-dependent memory; e.g., Godden & Baddeley, 1975). This same logic can be extended to a participant’s state (i.e., state-dependent memory) and mood (i.e., mood-dependent memory): Memory performance is greater when there is a match rather than a mismatch between a participant’s state or mood at study and test (e.g., Eich & Metcalfe, 1989; Goodwin, Powell, Bremer, Hoine, & Stern, 1969; Peters & McGee, 1982).

Thus, there are numerous variables that can be manipulated at study and/or test which can affect the encoding and retrieval of information, and ultimately memory for that information. One recent manipulation that has been found to affect memory is the production of information. Generally, memory performance tends to be greater for produced items (e.g., read aloud) compared to non-produced items (e.g., read silently); this memory phenomenon has been termed the production effect (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). While research exploring the production effect only began several years ago, the idea that saying items aloud may impact subsequent decisions regarding those items is much older.

1.2 HISTORY OF THE PRODUCTION EFFECT

The earliest recorded account of the production effect was found in a book written by Gates (1917). In his book, Gates (1917) reports an experiment in which participants provided a full report of all the functions (i.e., memory strategies) that they used while learning lists of items. Consistent with the production effect, nearly all of the participants reported that they found the pronunciation of information a particularly useful learning aid and specifically that pronunciation helped secure the “motor and auditory elements” of the words in memory (Gates, 1917). Thus, the idea that production results in a memory advantage has been around for nearly a century, if not longer.

While Gates (1917) provided subjective support for the production effect, Hopkins and Edwards (1972) were among the first researchers to provide objective support for the production effect. Hopkins and Edwards (1972; Hopkins, Boylan, & Lincoln, 1972) used the comparison of produced and non-produced items to test whether pronunciation of verbal units would enhance the perceived frequency of a word in both within- and between-subjects designs. To test this hypothesis, Hopkins and Edwards (1972) assigned participants to one of four instruction groups, wherein each group was presented with 100 study words, half of which were underlined: Participants were instructed to produce (a) only the underlined words, (b) only the non-underlined words, (c) neither the underlined nor non-underlined words, or (d) both the underlined and non-underlined words. Following the study phase, participants were tested using either forced-choice recognition (Experiment 1) or old/new recognition (Experiment 2). These manipulations permitted evaluation of the impact of production within subjects in cases (a) and (b) and between subjects by comparing performance in cases (c) and (d).

Hopkins and Edwards (1972) found that while memory was not greater for produced compared to non-produced words in between-subjects designs, memory was significantly greater for produced words than non-produced words in within-subjects designs. Furthermore, Hopkins and Edwards (1972) found that when memory performance was compared across the within- and between-subjects designs, there was no significant difference in memory performance for produced words; however, memory performance was significantly worse for non-produced words in the within-subjects design compared to the between-subjects design. Based on their findings, the authors concluded that the presence of the production effect is relative in two ways: 1) It requires the mixture of produced and non-produced words at encoding (the effect does not occur when participants are asked to produce all study words), and 2) it appears to be attributed to decreased memory performance for non-produced words, as opposed to increased memory performance for produced words —at least when using an old/new recognition test (the results were less conclusive for the forced-choice recognition test of Experiment 1).

Almost a decade later, Conway and Gathercole (1987; Gathercole & Conway, 1988) used a production paradigm similar to Hopkins and Edwards (1972) to investigate the effect of input modality —reading aloud, mouthing (i.e., making no vocalization, only a mouth movement) and reading silently —on incidental as opposed to intentional memory performance. Conway and Gathercole (1987) replicated the produced > non-produced difference in memory performance for words read aloud (memory performance for words that were mouthed was intermediate) which they interpreted as arising from differences in the sensory-perceptual activity involved in each condition; they assumed that non-sensory activity such as phonological processing was equivalent. Specifically, words that are read aloud involve both an articulatory/motor component (i.e., moving one's lips to produce the word); and an acoustic component (i.e., hearing oneself say the word), whereas words that are mouthed involve only an articulatory/motor component; words that are read silently involve neither an articulatory/motor nor acoustic component. The authors suggested that these differences in sensory-perceptual activity (i.e., articulatory and/or acoustic) at encoding give rise to differences in general and relational distinctiveness which, in turn, influence subsequent memory performance. Conway and Gathercole (1987) defined general distinctiveness as "…the degrees of freedom upon which an attribute may vary within a sensory domain…" and relational distinctiveness as "…the different [modality] conditions experienced within a particular episode" (p. 359). Essentially, general distinctiveness is the variability within forms of sensory-perceptual activity, such as the lip movements required for different words, and relational distinctiveness is similar to Hopkins and Edwards' (1972) concept of relativity in that the relative comparison of modality conditions at encoding is critical for a production effect to occur.

Gathercole and Conway (1988) extended their earlier research by examining whether the presence of both articulation/motor movement and acoustic information is necessary to observe a memory advantage or whether either alone is sufficient. Using a within-subjects design, Gathercole and Conway (1988) conducted four experiments, which differed according to the type of input conditions at study and included reading aloud, listening, mouthing, "seen" writing, "unseen" writing (i.e., where vision of what the participant wrote was hidden) or reading silently. Across these experiments, recognition performance was greatest for words in the read aloud condition and worst for words in the read silently condition with memory performance being intermediate for words in the listen condition (Experiment 1) and write seen condition (Experiment 3). In Experiment 4, there was no difference between words in the write unseen and read silently conditions. The core difference between the write seen and the write unseen conditions is that the write seen condition consists of both motor and visual elements, whereas the write unseen condition consists of only a motor element. Thus, while the results from

Experiment 4 suggests no effect of a pure motor element on the production effect, performance was numerically higher in the write unseen compared to the read silently condition. In addition, because their study used incidental rather than intentional learning, the results may not accurately represent the power of the production effect. That is, if participants had been producing the items with the specific aim of learning the words, there might have been a more robust production effect for the production conditions in Experiments 1 to 4. In a fifth experiment, Gathercole and Conway (1988) attempted to generalize their previous findings from a within-subjects design to a between-subjects design. Similar to Experiments 1-4, recognition performance was greatest for words in the read aloud condition compared to all other conditions, followed by memory performance for words in the listen condition; there were no differences in memory performance for the two (seen or unseen) written conditions, mouthed condition, and read silently condition.

In contrast to production studies using intentional learning (e.g., Hopkins & Edwards, 1972), under incidental learning conditions, memory performance was greatest for words in the read aloud condition compared to all other input conditions regardless of study design (within- versus between-subjects). This suggests that while both articulatory and acoustic information are not necessary to observe a memory advantage, when both articulatory and acoustic information are available, there is greater memory performance than when only one type of information is available (i.e., either articulatory or acoustic alone). In fact, in the between-subjects experiment, there was no memory advantage when only articulatory or motor information was available (i.e., the write and mouthed conditions).

The findings of these experiments also support the initial sensory-perceptual activity account proposed by Conway and Gathercole (1987): There is a relation between the variety of perceptual activity present at encoding and subsequent memory performance at test such that a greater variety of perceptual activity at encoding is associated with greater subsequent memory performance. Gathercole and Conway (1988) found that memory performance was greater when both articulatory and acoustic information was present in the read aloud condition followed by memory performance when only acoustic information was present in the listen and write seen conditions.

Ten years after the work of Gathercole and Conway (1988; Conway & Gathercole, 1987), MacDonald and MacLeod (1998) used a production manipulation at study (reading aloud versus reading silently) to explore the effects of attention on explicit and implicit memory tests. Their expectation was that participants would pay less attention to items read silently compared to items read aloud. In their first experiment, participants studied a word list while either reading each item aloud or saying “pass” instead, depending upon whether the word was red or white; their second experiment instead replaced the “pass” condition with a more typical read silently condition. In both experiments, memory performance was tested using a rapid reading test intended to index implicit memory (MacLeod, 1996; MacLeod & Masson, 1997) followed by an old/new recognition test intended to index explicit memory. In each experiment, only the explicit memory test demonstrated an advantage for words read aloud compared to words read silently, suggesting that the production effect does not generalize to implicit memory tests – at least insofar as speeded reading is concerned (see also Hourihan & MacLeod, 2008; MacLeod et al., 2010).

While the production effect has been incorporated as a manipulation in past research to study other theories and phenomena, such as word frequency theory (Hopkins & Edwards, 1972), long-term retention (Conway & Gathercole, 1987; Gathercole & Conway, 1988), and attention (MacDonald & MacLeod, 1998), it is only recently that the production effect has been widely researched as a memory phenomenon in its own right (see MacLeod et al., 2010). Although the study items are usually lexical in nature (e.g., single words, word pairs, sentences, paragraphs; see Ozubko, Hourihan, & MacLeod, 2012), other stimuli such as non-words (MacLeod et al., 2010; Experiment 6), and pictures (Fawcett, Quinlan, & Taylor, 2012; Richler, Palmeri, & Gauthier, 2013) also result in significant production effects. In addition, reading or naming the study item aloud tends to be the most common form of production; however, other methods have included spelling, writing, and typing (Forrin, MacLeod, & Ozubko, 2012; Richler et al., 2013); mouthing (Castel, Rhodes, & Friedman, 2013; Fawcett et al., 2012; Forrin et al., 2012; MacLeod et al., 2010); whispering (Forrin et al., 2012; see also Castel, 2009, and Castel et al., 2013); and listening to the auditory presentation of words (Forrin & MacLeod, 2016). The production effect has been most often measured using yes/no recognition, but some studies have instead included fill-in-the-blanks (Ozubko et al., 2012), forced-choice recognition (MacLeod et al., 2010), and free recall (e.g., Jones & Pyc, 2014; Jonker, Levene, & MacLeod, 2013), often following a short delay (Hourihan & MacLeod, 2008; MacLeod et al., 2010) but sometimes following a long delay (Ozubko et al., 2012) between study and test.

As detailed above, earlier studies used production as *a study manipulation* to examine other effects (e.g., frequency effect, attention), whereas more recent studies have used production *to directly examine* the production effect as a memory phenomenon. The production effect has now been widely researched as a memory phenomenon (see Bodner & MacLeod, 2016; MacLeod & Bodner, 2017) and has been shown to be robust across a variety of study manipulations including various types of stimuli, production, and memory tests. At present, research suggests that reading aloud is the most effective form of production, resulting in greater memory performance than other forms of production, such as spelling, writing, typing, mouthing, and whispering (e.g., Castel 2009; Castel et al., 2013; Fawcett et al., 2012; Forrin et al., 2012; MacLeod et al., 2010).

1.3 A REVIEW OF THEORETICAL PERSPECTIVES

Many researchers (e.g., Fawcett et al., 2012; Forrin et al., 2012; Forrin & MacLeod, 2016; Ozubko & MacLeod, 2010) have argued that the production effect is best explained by a distinctiveness account, which assumes that producing an item results in a relatively more “distinct” memory trace, making produced items easier to retrieve at test compared to non-produced items (see Schmidt, 1991, and more recently, Hunt, 2006, for a review of distinctiveness).

In particular, MacLeod et al. (2010) and Ozubko and MacLeod (2010) have proposed that such a distinctiveness account assumes that compared to non-produced items, produced items have at least one distinct element that is encoded at study, which serves as a retrieval cue to guide memory performance at test. For example, compared to reading silently, reading aloud consists of two additional distinct elements: articulation and audition, which are encoded at study. At test, participants can use these two distinct elements to decide whether an item was studied: If participants remember saying the word aloud and/or hearing themselves say the word aloud, they can use that information to decide that the item was presented at study. Thus, in the production paradigm, participants may use a memory heuristic (see Schacter, Israel, & Racine, 1999) whereby distinct elements that were encoded at study are used to guide retrieval at test. Similar to reading aloud, other forms of production, such as spelling, typing, writing, and mouthing (see Forrin et al., 2012) are presumed to involve an additional distinct motor element that is not present while reading silently.

As already noted, the production effect is consistently found in individual studies that use a within-subjects design, but more rarely found in individual studies that use a between-subjects design (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010). The lack of significant production effect in between-subjects studies has been taken as support for a distinctiveness account and suggests that a comparison of the relative distinctiveness across studied items is critical (although see Gathercole & Conway, 1988). In particular, MacLeod et al. (2010) argued that without a “backdrop” of silently read study items such as provided only in a within-subjects presentation, the distinct elements associated with reading aloud are no longer useful as retrieval cues at test. Moreover, Ozubko and MacLeod (2010) provided further evidence for the distinctiveness account by demonstrating that the production effect was eliminated when the distractor items to be used at test were read aloud preceding the study phase; the production effect was still observed when the distractor items to be used at test were read silently preceding the study phase. A distinctiveness account would predict that reading distractor items aloud functions to disrupt the production effect because the distinctive features associated with reading aloud at study are no longer useful retrieval cues at test if all test items have been read aloud (Ozubko & MacLeod, 2010; although see Bodner & Taikh, 2012, for an alternative explanation).

Further support for a distinctiveness account comes from the relation between the potential number and strength of distinct elements and subsequent memory performance. For example, Fawcett et al. (2012) demonstrated an interaction between the production effect and the picture superiority effect, which they suggested arose because producing a word resulted in at least two distinct elements (articulation and audition), whereas producing a picture resulted in at least three distinct elements (articulation, audition, and visual detail). As a result, they observed an interaction for *d’* scores between production and picture presentation such that the production effect was much larger for pictures than words.

Likewise, Forrin et al. (2012) provided a demonstration of this principle by revealing that different types of production are associated with differences in the magnitude of the production effect. For instance, while they found a significant production effect for non-vocal forms of production, such as writing and mouthing, the size of that production effect was smaller compared to reading items aloud. Similar to Fawcett et al. (2012), these researchers explained these results by suggesting that the greater the number of potential distinct elements at encoding, the larger the production effect. Compared to reading silently, both writing and mouthing involve an additional distinct motor element (the hand movement associated with writing and the lip movement associated with mouthing), whereas reading aloud involves not only a distinct motor element (i.e., the lip movement associated with reading aloud) but also the additional distinct element of audition (i.e., hearing the word being said aloud).

Interestingly, Forrin et al. (2012) also examined the size of the production effect for whispering versus reading aloud in a normal voice and found that the production effect was smaller when participants whispered items compared to when they read items aloud. This suggests that intensity may also be an important distinct element that is encoded at study and influences subsequent memory performance at test. Indeed, Castel (2009) found a linear trend for the production effect to be greater when items were read aloud in a loud voice compared to when items were whispered (see also Castel et al., 2013). Similar to qualitative distinct elements (e.g., articulation and audition), perhaps quantitative distinct elements such as intensity provides a greater initial capture of attention and early processing at encoding, which affects the subsequent information available to be used as retrieval cues at test. Thus, based on recent findings in the literature (Castel, 2009; Castel et al., 2013; Fawcett et al., 2012; Forrin et al., 2012), there appears to be an association between the amount of distinct information available at encoding and the size of the subsequent production effect at test. Together, these findings support a distinctiveness account by demonstrating that distinct elements encoded at study may be used to guide effective retrieval at test.

While the majority of research on the production effect supports a distinctiveness account, some theorists have considered whether the effect might instead be attributable to differences in the general "strength" or elaboration of the relevant trace within memory (e.g., Bodner & Taikh, 2012; Ozubko & MacLeod, 2010). For example, an encoding-centric or strength-based account might assume that produced items are encoded more elaborately than non-produced items, improving performance during an appropriate retrieval task (also see Craik, 2002; Craik & Lockhart, 1972). Importantly, such an account need not make any assumptions concerning the strategies that participants employ during that task. Rather, the produced items are merely characterized by a "stronger" or more elaborate representation with greater retrieval potential than non-produced items. Such a theoretical account is conceptually similar to a levels of processing account (Craik & Lockhart, 1972) wherein memory tends to be greater for items that are processed at a deeper semantic and/or cognitive level compared to those that are not. That is, there is a more persistent and durable memory representation for deeply or elaborately processed items.

Recently, Fawcett (2013) found that when all between-subjects production effect studies are aggregated and analyzed using meta-analytic techniques, there is a small but significant, production effect. At first glance, this finding seems to be incompatible with a distinctiveness account, but consistent with a strength-based account. However, as suggested by Ozubko, Major, and MacLeod (2014), it may be that when participants read an entire list of words aloud in a between-subjects design, the act of reading aloud is no longer considered distinct and thus, they do not use (or are less likely to use) that information at test.

Fawcett (2013) argued that the comparison of the magnitude of the production effect for within- and between-subjects designs might provide clarification regarding the mechanisms underlying the production effect. He speculated that the finding of a less robust between-subject production effect could suggest that the within-subject production effect involves processes at both encoding and retrieval whereas the between-subject production effect involves only one of those processes. For example, it is possible that production results in a more elaborate or familiar (i.e., "stronger") memory trace regardless of study design, producing a benefit to memory even in between-subjects designs; however, when production is manipulated within subjects, participants might also allocate additional attention and distinct processing to the produced versus non-produced items and use the distinct processing to guide performance at test. Consistent with the rationale provided by Fawcett (2013), recent research found that the production effect in within-subjects designs is much larger than the production effect in between-subjects designs (magnitude of 20% versus 4%; see Bodner, Taikh, & Fawcett, 2014, Experiment 1; also see Forrin, Groot, & MacLeod, 2016). Thus, in accordance with Fawcett (2013) and Ozubko et al. (2013), the presence of a small production effect in between-subjects designs is not entirely incompatible with a distinctiveness account for within-subjects study designs. It may simply reflect the inconsistent (or lack of) use of a distinctiveness heuristic in between-subjects designs. Furthermore, the inconsistent presence of a production effect in between-subjects designs may reflect multiple processes underlying the production effect, which is largely dependent upon study design (i.e., between versus within-subjects designs).

As evident from this summary of the literature, the findings from studies in the field are mixed: Some studies support a distinctiveness account (Fawcett et al., 2012; Forrin et al., 2012; Forrin et al., 2016; Ozubko et al., 2012; Ozubko & MacLeod, 2010) and other studies hint at a strength-based or encoding centric account (Bodner & Taikh, 2012) or some combination of these processes (Fawcett, 2013; Fawcett & Ozubko, 2016; Jamieson, Mewhort, & Hockley, 2016; Jonker et al., 2013; Ozubko, Gopie, & MacLeod, 2012; Taikh & Bodner, 2016).

1.4 MUSIC AND MEMORY

As evident in the production effect literature, reading aloud and other forms of production (e.g., spelling, writing, mouthing, and whispering; see

Forrin et al., 2012) have been shown to be very effective mnemonics — regardless of the underlying mechanism(s). Perhaps it should not be surprising that music is also considered to be a unique and powerful memory tool. Music is an effective memory tool because it provides melodic, rhythmic, and temporal organization of information, allowing for information to be easily organized into more manageable units (Thaut, Peterson, & McIntosh, 2005). Thaut et al. (2005; pp. 252) explain that using a musical template or song (i.e., rhythm, melody, pitch/timbre) for verbal learning induces greater "synchrony in learning related networks" thereby producing "more stable neural traces for long-term memory," and consequently increasing access to those memories. Music as a mnemonic has been applied to many clinical populations. For example, music has been shown to be an effective tool for learning nonmusical materials in individuals with learning and developmental disorders (Claussen & Thaut, 1997; Gfeller, 1983; Wolfe & Hom, 1993). Structured music listening has also been shown to enhance a wide range of cognitive functioning in individuals with autism spectrum disorder (Bettison, 1996).

In addition, researchers have found that musical memories are preserved in individuals with neurological memory disorders (Baur, Uttner, Ilmberger, Fesl, & Mai, 2000; Haslam & Cook, 2002). For example, music can enhance the recall of autobiographical memories in individuals with Alzheimer's disease (e.g., El Haj, Facotti, & Allain, 2012; Foster & Valentine, 2001; Irish, Cunningham, Walsh, Coakly, Lawlor, Robertson, et al., 2006). In fact, compared to memories recalled without music, those recalled in the presence of music tend to be more specific, to contain more emotional content, and also to be retrieved more quickly than those in the absence of music (El Haj et al., 2012). Recently, Oostendorp and Montel (2014) conducted a study wherein individuals diagnosed with severe Alzheimer's disease were given an encoding task followed by a recall task. During the encoding task, participants were given two word lists and assigned to one of two conditions: Sing or read aloud. Participants recalled significantly more words in the sing condition compared to the read aloud condition. These researchers attributed their findings to three possible explanations: 1) a sparing of brain areas associated with music in individuals with Alzheimer's disease; 2) the use of a musical template to facilitate verbal learning via greater neural synchrony in learning-related networks (see Thaut et al., 2005, for further discussion); and/or 3) a positive emotional reaction to the music in areas of the brain (i.e., hippocampus, prefrontal cortex) that play a major role in memory encoding and retrieval strategies (Oostendorp & Montel, 2014). Additional explanations are suggested by the fact that music and singing have been hypothesized to have an effect on arousal, resulting in improved memory due to attention (Simmons-Stern, Budson, & Ally, 2010). Thus, overall, the research just described suggests that music and singing are powerful mnemonics that can be easily implemented and that improve memory performance across several populations, including patients who have significant memory deficits (i.e., individuals who have Alzheimer's disease, developmental delays, and learning disorders).

1.5 CHAPTER SUMMARY, RATIONALE, AND CURRENT EXPERIMENTS

In summary, over the past several years, the production effect has become a widely researched memory phenomenon that has been studied using a broad range of study and test characteristics. The effect occurs across a number of types of production including reading aloud, whispering, mouthing, spelling, writing, and typing (see Forrin et al., 2012), as well as for both picture and word stimuli (Fawcett et al., 2012). While the effect consistently occurs in within-subjects study designs, it is not commonly observed in between-subjects designs (MacLeod et al., 2010; although see Fawcett, 2013) and when it does occur in between-subjects designs, its magnitude is smaller compared to within-subjects designs. Despite alternative explanations, the production effect is typically explained by a distinctiveness account. This account suggests that the memory advantage for produced versus non-produced study items occurs because produced items are distinct in relation to a backdrop of non-produced study items. In other words, produced items tend to be distinct in the context of non-produced items. In addition, researchers have hypothesized that the potential number and strength of distinct elements is related to subsequent memory performance and thus to the magnitude of the production effect (e.g., Castel, 2009; Castel et al., 2013; Forrin et al., 2012).

Based on the rationale of Forrin and colleagues (2012), it follows that any form of production that has a greater number or strength of distinct elements than reading aloud should result in greater subsequent memory performance. Compared to reading aloud, singing contains melody and rhythm as well as dynamic elements, such as intensity, pitch, and timbre (Roederer, 2008). Thus, singing provides several qualitative and quantitative differences that could be used to distinguish items that are sung from those that are read aloud. In addition, as outlined above, music and singing are efficacious mnemonics that have been found to improve memory performance in several clinical populations including individuals who have Alzheimer's disease (Oostendorp & Montel, 2014) and individuals who have developmental delays or learning disorders (Claussen & Thaut, 1997; Gfeller, 1983; Wolfe & Hom, 1993). Thus, based on research in clinical populations, it appears as though singing has the potential to be a more powerful form of production than reading aloud.

As such, there were three main goals of the present dissertation: 1) to use the production paradigm as a tool to examine the effects of singing on memory performance in a nonclinical population; 2) to expand our knowledge of the production effect by incorporating a form of production (i.e., singing) that has been hypothesized to be a very powerful mnemonic; and 3) to explore the mechanism(s) underlying any observed effects of singing on memory performance. Overall, incorporating singing into the production paradigm will further our understanding of the mechanisms through which singing influences memory in a nonclinical population, as well as our understanding of the production effect and its underlying mechanisms.

In Experiments 1 to 3 of the present dissertation, we examined whether other forms of production, such as reading aloud in a loud voice and singing were more effective forms of production compared to reading aloud in a normal voice. Indeed, both reading aloud loudly and singing produced a greater production effect compared to reading aloud; memory performance was also greater for sing items versus read aloud loudly items. Although the results of Experiments 1-3 are compatible with a distinctiveness account of the production effect, the remaining experiments were designed to rule out other possible explanations.

The primary goal of Experiment 4 was to examine whether the greater production effect for singing versus reading aloud could be attributed to the simple fact that singing may be an unusual or bizarre action for many people. Using participants who were experienced singers, Experiment 4 replicated the findings of Experiments 2 and 3: Singing produced a greater production effect compared to reading aloud. Experiment 5 explored the possibility that the act of singing may inherently require a longer production duration than reading aloud. If a longer production duration for sing items is driving the greater production effect for singing, then explicitly instructing participants to reduce the production duration for sing items and to increase the production duration for read aloud items should function to reduce or eliminate the greater production effect for sing versus read aloud. In contrast, the results from Experiments 2, 3, and 4 were replicated: Singing produced a greater production effect than reading aloud.

Experiment 6 used a between-subjects study design to determine whether differences in elaborate encoding and strength of memory representations could account for the greater production effect for singing versus reading aloud. We hypothesized that if singing produces a greater production effect than reading aloud due to increased processing, elaborate encoding, and strength of memory representations, we would find a significant production effect for sing items compared to read aloud and read silently items, even in a between-subjects design. Our findings are similar to those from other forms of production (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010; although see Fawcett, 2013, for review): Singing did not yield a between-subjects production effect, which is inconsistent with a pure strength-based account.

In sum, the experiments detailed in this dissertation suggest that there are more effective forms of production than reading aloud. Across five within-subjects experiments, memory performance was significantly greater for sing versus read aloud and read silently items. Importantly, as indicated by the greater production effect for singing versus reading aloud in experienced singers and under reduced processing time, as well as the non-significant between-subjects production effect, the greater memory performance for singing versus reading aloud is most compatible with a distinctiveness account of the production effect.

CHAPTER 2: EXTENDING THE BOUNDARIES OF THE PRODUCTION EFFECT

The following chapter has been previously published as:

Quinlan, C. K. & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory, 21*, 904-915.

Although Dr. Tracy L. Taylor was a co-author for this manuscript, Chelsea K. Quinlan was the primary contributing author to this manuscript; she designed, programmed, conducted, and analyzed and interpreted the data from all experiments. Also, she produced the first initial draft of this manuscript as well as all major revisions. See Appendix A for copyright permissions.

2.1 ABSTRACT

The production effect is the finding that subsequent memory is better for words that are produced than for words that are not produced. Whereas the current literature demonstrates that reading aloud is the most effective form of production, the distinctiveness account used to explain the production effect predicts that there is nothing special about reading aloud per se: Other forms of vocal production that include an additional distinct element should produce even greater subsequent memory benefits than reading aloud. To test this, we presented participants with study words that they were instructed to read aloud loudly, read aloud normally, or read silently (Experiment 1); sing, read aloud, or read silently (Experiment 2); and sing, read aloud loudly, read aloud, or read silently (Experiment 3). We observed that both reading items aloud loudly (Experiment 1; 3) and singing items (Experiment 2; 3) at study resulted in greater subsequent recognition than did reading items aloud in a normal voice; furthermore, singing had a larger memory benefit than did reading aloud loudly (Experiment 3). Our findings support the distinctiveness hypothesis by demonstrating that there are other forms of production, such as singing and reading aloud loudly, that have a more pronounced effect on memory than reading aloud because they add further distinctive features.

2.2 INTRODUCTION

In a study conducted by MacLeod and colleagues (2010), participants were presented with a series of study items, half of which they read aloud and half of which they read silently. MacLeod et al. (2010) found that subsequent memory was significantly greater for items that were read aloud compared to items that were read silently. This difference in memory performance for produced and non-produced items is now known as the production effect (although see Conway & Gatherole, 1987; Hopkins & Edwards, 1972, for earlier research in this field).

The production effect is most commonly explained by a distinctiveness account (see Ozubko & MacLeod, 2010). This account suggests that, compared to items that are not produced, items that are produced have *at least* one additional distinct element that is encoded at study and that can support subsequent retrieval at test. For instance, items that are read aloud versus silently include two distinct elements: articulation and audition. At test, if participants remember producing an item aloud, they can use that information to decide that the item was presented at study; however, if participants do not have a memory of producing that item aloud (i.e., as with reading a study word silently), this information is insufficient to decide whether the item was presented at study because the foil items were also not produced. Supporting a distinctiveness account, when participants are asked to read aloud the distractor items, the distinct element of having read aloud the study items is not useful for discriminating studied from distractor items in a list discrimination task and the usual memory advantage for items read aloud versus items read silently disappears (Ozubko & MacLeod, 2010). Thus, it is *relative* distinctiveness of the study items that is the critical determinant of the production effect. Indeed, while a recent meta-analysis suggests a moderate effect of production in between-subjects designs (Fawcett, 2013), it typically does not occur (or is less robust) in between-subjects designs (MacLeod et al., 2010; also see Dodson & Schacter, 2001; Hopkins & Edwards, 1972) and in pure list study designs (Ozubko & MacLeod, 2010), precisely because the distinctiveness of study items must be evaluated relative to other study items to be most effective.

While the production effect is commonly studied by comparing memory performance for items read aloud and items read silently, Forrin and colleagues (2012) extended the earlier work of MacLeod et al. (2010) by examining other forms of production. In a three level mixed-list design, Forrin et al. (2012) found that reading aloud at study resulted in better subsequent memory than three other forms of production: writing (Experiment 2A), mouthing (Experiment 2B), and whispering (Experiment 2C) —all of which produced better memory than reading silently but worse memory than reading aloud. Forrin et al. (2012) argued that, relative to items that are read silently, memory performance is better for items that are written and mouthed because writing and mouthing each involve a distinct motor element (i.e., manual movement and articulation, respectively); there is an extra benefit for reading aloud because it involves the additional distinct element of audition. The difference in memory performance for reading aloud in a normal voice and whispering suggests dependence of the production effect on the *intensity* of the distinct element available in the production. Similarly, Castel (2009) found a trend (although non-significant) for a greater production effect for items read aloud loudly compared to items read aloud quietly (see also Castel et al., 2013). It thus appears that cues to relative distinctiveness can be defined both qualitatively (e.g., motor, motor + audition) and quantitatively (e.g., the intensity of audition), with the effect on subsequent memory depending on the total number of available cues.

While Forrin et al. (2012) showed that the magnitude of the production effect is larger for reading aloud compared to other forms of production and argued that this is related to the number and strength of the distinct elements available in the type of production, it nevertheless remains possible that there is something special about reading aloud *per se* that makes it a stronger cue to relative distinctiveness than other forms of production. We ruled out this possibility by examining the effects of two additional vocal productions on subsequent memory: reading aloud loudly and singing. On the premise that reading aloud loudly includes an additional distinctive element (increased intensity of vocalization and audition) over reading aloud in a normal voice, the distinctiveness account predicts better subsequent memory for items read aloud loudly than for items read aloud in a normal voice. Indeed, this would fit with Forrin et al.' s (2012) demonstration of a larger production effect for reading aloud in a normal voice compared to whispering.

Extending this reasoning, singing aloud may include dynamic elements of intensity, pitch, and/or timbre (Roederer, 2008), thus providing both qualitative and quantitative differences to distinguish sung items from those read aloud loudly or in a normal voice. A distinctiveness framework would predict greater subsequent memory for singing aloud relative to reading aloud either loudly or in a normal voice. If these predictions are borne out, they will make it very clear that there is nothing special about reading aloud *per se* that results in a large subsequent memory advantage, but that the critical factor is, as MacLeod and colleagues have argued (Forrin et al., 2012; MacLeod et al., 2010), the number of distinct elements encoded at study that serve as cues to relative distinctiveness.

2.3 EXPERIMENT 1

The goal of Experiment 1 was to determine whether reading items aloud loudly at study would result in better subsequent recognition memory performance than reading items aloud in a normal voice or silently. In the study phase, participants viewed a series of words, one at a time. On each study trial, the word was printed in one of three colors: red, blue, or white. Each color corresponded with a specific instruction: Read Aloud Loudly, Read Aloud (in a normal voice), or Read Silently. Immediately following the study phase, we asked participants to complete a yes/no recognition test, which included all of the study items as well as an equal number of foil items not presented at study. We used the yes/no recognition test to calculate hit rates as well as foil false alarm rates for each of the three within-subjects conditions (Read Aloud Loudly, Read Aloud, Read Silently). By calculating both separate hit rates as well as separate foil false alarm rates, we were able to calculate the nonparametric measure of *A'*, which we used to indicate the signal strength or sensitivity/discriminability of memory performance in the three conditions (Read Aloud Loudly; Read Aloud; Read Silently; see Donaldson, 1996). The distinctiveness account predicts that reading aloud loudly will result in larger A' scores than reading aloud in a normal voice, and that reading aloud in a normal voice will result in larger A' scores than reading silently (the standard production advantage; MacLeod et al., 2010).

2.3.1 Method

Participants

Fourteen undergraduate students (4 males, 10 females) participated in this experiment in exchange for credit toward their grade in an eligible Psychology class at Dalhousie University. The experiment was run in one session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language.

Stimuli and Apparatus

This experiment was run using PsyScope 5.1.2 (Cohen, MacWhinney, Flatt, & Provost, 1993) loaded on a 24” iMac computer running Mac OSX Leopard, version 10.5. All responses were recorded on a standard Macintosh Universal Serial Bus keyboard. All text was presented against a uniform black background in Time New Roman size 42 font.

Two hundred and forty words were selected from the Paivio, Yuille, and Madigan (www.math.yorku.ca/SCS/Online/paivio/) word generator. The words were all nouns, three to seven letters in length with a mean of 5.20 letters and 1.55 syllables. The words had a mean Kucera-Francis (1967) word frequency of 77.31, a mean imagery rating of 5.47, a mean concreteness rating of 5.34, and a mean meaningfulness rating of 6.45. Both the study words and the yes/no recognition words were printed in a red, blue, or white colored font, which represented three conditions: Read Aloud Loudly, Read Aloud, and Read Silently. For the yes/no recognition phase, words were also printed in red, blue, or white colored font, which provided us with three separate foil false alarm rates that were used to calculate A' for the study lists presented in the corresponding color (e.g., if red signaled a Read Aloud Loudly instruction at study, the foil items presented in red provided the false alarm rate for calculations made in the Read Aloud Loudly condition).

Prior to running each participant, custom software randomly distributed the 240 words into three study lists (Read Aloud Loudly, Read Aloud, Read Silently), each consisting of 40 words, and three foil lists (i.e., items not presented at study), each consisting of 40 words.

Procedure

Before beginning the experiment, participants were given verbal instructions, which were re-iterated on the computer monitor (See Appendix B for an example of experiment instructions). The experimenter told participants that in the study phase they would see a series of words printed in red, blue, or white. One third of participants were told to read aloud loudly the words in red, read aloud normally the words in blue, and read silently the words in white; one third of participants were told to read aloud loudly the words in blue, read aloud normally the words in white, and read silently the words in red; and one third of participants were told to read aloud loudly the words in white, read aloud normally the words in red, and read silently the words in blue. The experimenter told participants that they would be required to complete a memory test following the presentation of all study words.

Before beginning the experiment proper, the experimenter told participants that they would be presented with a familiarization phase and a practice phase that were designed to ensure that they were comfortable with the three instruction conditions (Read Aloud Loudly, Read Aloud, Read Silently). For the Read Aloud Loudly condition, participants were instructed to read more loudly than they would in normal day-to-day interaction (e.g., so that a friend on the other side of room could hear them). The experimenter remained in the room with participants until the end of the study phase, to ensure that participants followed the instructions for the three conditions (Read Aloud Loudly, Read Aloud, Read Silently).

Familiarization phase. Prior to beginning the study phase, participants were presented with 15 trials. These trials were designed to familiarize participants with the three font colors (red, blue, white) and their associated instruction (Read Aloud Loudly, Read Aloud, Read Silently). Five trials of each color and its associated condition were intermixed randomly. On each familiarization trial, a blank screen was presented for 500 ms followed by the verbal descriptor of the color along with its associated condition (e.g., ‘RED-Read Aloud Loudly’) at centre for 2000 ms. Both the names of the color as well as the associated condition were printed in the indicated color (e.g., ‘RED-Read Aloud Loudly’ was printed in red).

Practice phase. Immediately following the familiarization phase, participants completed a practice phase. On each practice trial, a blank screen was presented for 500 ms followed by the word 'banana' at centre for 2000 ms. The word 'banana' was printed in one of three colors (red, blue, white), which corresponded to one of three conditions (Read Aloud Loudly, Read Aloud, Read Silently). Five trials of each condition were intermixed randomly to produce a total of 15 practice trials. These practice trials were identical to those in the study phase with the exception that 'banana' was the only word presented. This phase was designed to give participants practice with the three conditions (Read Aloud Loudly, Read Aloud, Read Silently).

Study phase. Immediately following the last trial in the practice phase, the study phase trials began. The study phase consisted of a total of 120 trials. There were 40 trials in each of the three conditions (Read Aloud Loudly, Read Aloud, Read Silently), which were intermixed randomly.

Each study phase trial began with a blank screen for 500 ms followed by a word in the centre for 2000 ms. Each word was selected randomly without replacement from the Read Aloud Loudly, Read Aloud, and Read Silently study lists. The total duration of each study trial was 2500 ms.

Recognition phase. Upon completing the study phase, participants began the recognition phase. The recognition phase consisted of a self-paced yes/no recognition test. At the beginning of the recognition phase, instructions were presented at the top of the computer screen. Participants were instructed to press the 'y' key if they recognized the word from the study trials and to press the 'n' key if they did not recognize the word from the study trials (i.e., a foil word). All responses could be self-corrected using the backspace key and were submitted by pressing the space bar. The next recognition trial began after each response was submitted.

Because we were interested in using a signal detection approach to calculate a measure of sensitivity (*A'* scores), it was necessary to estimate separate foil false alarm rates for each condition (see Macmillan & Creelman, 2005). To calculate separate foil false alarm rates for each condition, we maintained the colour coding from the study phase such that items that were presented in red were tested in red; items that were presented in blue were tested in blue; and items that were presented in white were tested in white. We also equally distributed the colour assigned to the foil items such that an equal number of foil items were presented in red, blue, and white coloured fonts. As mentioned above, the font colours and production conditions were counterbalanced across participants.

As such, on each recognition trial, a word printed in red, blue, or white colored font was presented at the centre of the computer monitor until a response was made. In total, there were 240 recognition trials: the 120 words presented in the study phase and 120 foil words; these were intermixed randomly. The 120 items from the three study lists were printed in the same color as they were in the study phase: 40 study items were printed in red, 40 study items were printed in blue, and 40 study items were printed in white. The 120 items from the three foil lists were printed in colors that corresponded to the study phase conditions:

40 foil items were printed in red, 40 foil items were printed in blue, and 40 foil items were printed in white. Although the items in the yes/no recognition phase were printed in different colors, participants were not instructed to perform the condition (Read Aloud Loudly, Read Aloud, Read Silently) corresponding to the color of the item (red, blue, white).

2.3.2 Results

A hit was defined as a ‘yes’ response to studied words from the Read Aloud Loudly, Read Aloud, and Read Silently conditions; a false alarm was defined as a ‘yes’ response to unstudied foil words. The false alarms were classified according to the meaning of the color-coding at study (e.g., if a participant responded "yes" to the recognition of a foil word printed in red and red had signaled the Read Aloud Loudly condition at study, the foil response was considered to be a false alarm for the Read Aloud Loudly condition). The mean proportions of hits and foil false alarms are shown in Table 1.

Table 1
Means (and standard deviations) for the proportion of hits as well as foil false alarms as a function of condition (Read Aloud Loudly, Read Aloud, Read Silently).

Instruction Condition	**Hits**	**Foil False Alarms**
Read Aloud Loudly	.783 (.094)	.253 (.161)
Read Aloud	.657 (.112)	.247 (.174)
Read Silently	.506 (.149)	.204 (.146)

Sensitivity (*A’*) was calculated on a subject-by-subject basis and analyzed in a one-way ANOVA with condition (Read Aloud Loudly, Read

Aloud, Read Silently) as a within-subjects factor. Lower *A'* values represent lower sensitivity and higher *A'* values represent greater sensitivity (a value of .50 represents chance performance). As seen in Figure 1, there was a significant effect of condition, $F(2, 26)=26.362$, $MSe=.001$, $p<.001$, ($\eta^2=.670$). Planned contrasts showed that the *A'* scores were significantly greater in the Read Aloud Loudly condition ($M=.845$, $SD=.066$) than in both the Read Aloud condition ($M=.793$, $SD=.082$; $t(13)=3.928$, $p=.002$), and the Read Silently condition, ($M=.749$, $SD=.077$; $t(13)=7.907$, $p<.001$). In turn, *A'* scores were greater in the Read Aloud condition than in the Read Silently condition, $t(13)=3.086$, $p=.009$ (see Appendix C for additional analyses).

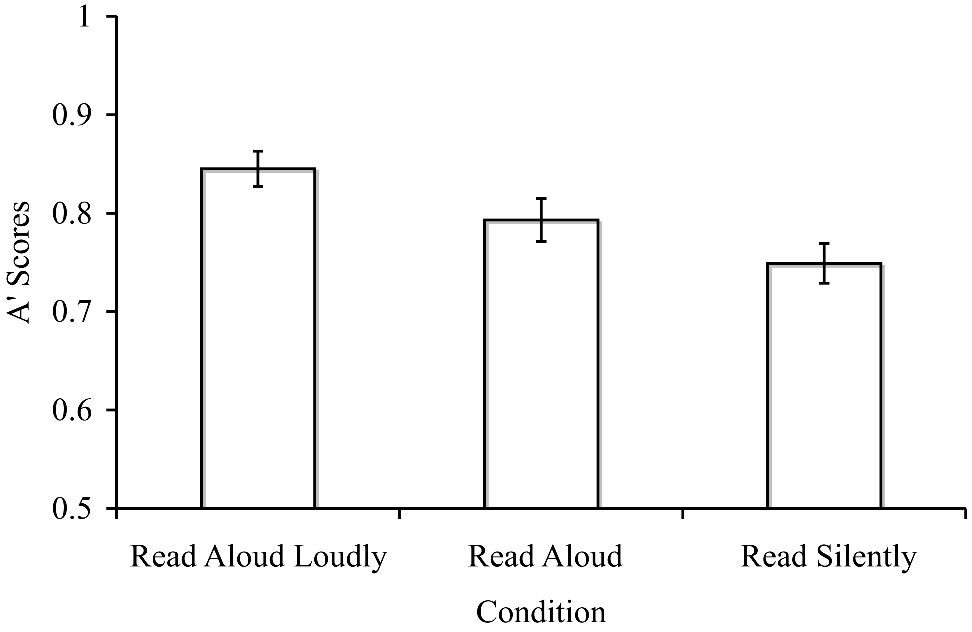


Figure 1. Experiment 1: The *A'* scores on the recognition test as a function of condition (Read Aloud Loudly, Read Aloud, Read Silently); error bars represent one standard error.

2.3.3 Discussion

The purpose of this experiment was to determine whether reading items aloud loudly at study would result in even better subsequent memory performance than reading items aloud in a normal voice. Indeed, while there was greater item discriminability for words that were read aloud compared to words that were read silently, there was even greater item discriminability for words that were read aloud loudly versus words that were read aloud in a normal voice. In the *A'* scores, the production effect was .096 for items that were read aloud loudly and .044 for items that were read aloud, reflecting a .052 advantage in subsequent memory due to the intensity of the production (i.e., read aloud loudly versus in a normal voice). These results demonstrate that there is nothing special about reading aloud *per se* that results in a large production effect. Instead, the magnitude of the production effect is graded according to the intensity of the verbal production, with a larger effect for items read aloud loudly versus in a normal voice (Experiment 1) and a larger effect for items read aloud in a normal voice versus whispered (Forrin et al., 2012).

2.4 EXPERIMENT 2

The goal of Experiment 2 was to determine whether singing at study would result in better subsequent memory performance than reading aloud. We essentially replicated the methods of Experiment 1, except that the instruction to read aloud loudly was replaced by the instruction to sing. We reasoned that items that were sung aloud likely included a number of distinct elements over and above articulation and audition, including dynamic elements of intensity, pitch, and/or timbre (Roederer, 2008). Assuming that the production with the greater number of distinct elements will result in better subsequent memory, we thus predicted that there would be greater memory for items sung than for items read aloud, as well as greater memory for items read aloud than for items read silently.

2.4.1 Method

Participants

Twenty undergraduate students (6 males, 14 females) participated in this experiment in exchange for credit toward their grade in an eligible Psychology class at Dalhousie University. The experiment was run in one session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language.

Stimuli and Apparatus

The stimuli and apparatus were identical to Experiment 1.

Procedure

The general procedure was identical to Experiment 1 with the exception that the Sing condition replaced the Read Aloud Loudly condition. This resulted in three instruction conditions: Sing, Read Aloud, and Read Silently. For the Sing condition, participants were instructed to sing as they would in any other context (e.g., in the car, in the shower) and thus the singing strategies used by participants varied depending on their individual style.

2.4.2 Results

The mean proportions of hits and foil false alarms were calculated in the same manner as described in Experiment 1 and are shown in Table 2.

Table 2
Means (and standard deviations) for the proportion of hits as well as foil false alarms as a function of condition (Sing, Read Aloud, Read Silently).

Instruction Condition	**Hits**	**Foil False Alarms**
Sing	.651 (.124)	.193 (.126)
Read Aloud	.555 (.154)	.217 (.135)
Read Silently	.396 (.187)	.167 (.143)

Sensitivity (*A'*) was calculated on a subject-by-subject basis and analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. As seen in Figure 2, there was a significant effect of condition, *F*(2, 38)=19.28, *MSe*=.003, *p*<.001, (η^2=.504). Planned contrasts showed that the *A'* scores were significantly greater in the Sing condition (*M*=.823, *SD*=.051) than in both the Read Aloud condition (*M*=.769, *SD*=.073; t(19)=4.800, *p*<.001), and the Read Silently condition (*M*=.723, *SD*=.083; *t*(19)=5.817, *p*<.001). In turn, the *A'* scores were greater in the Read Aloud condition than in the Read Silently condition, *t*(19)=2.271, *p*=.035 (see Appendix D for additional analyses).

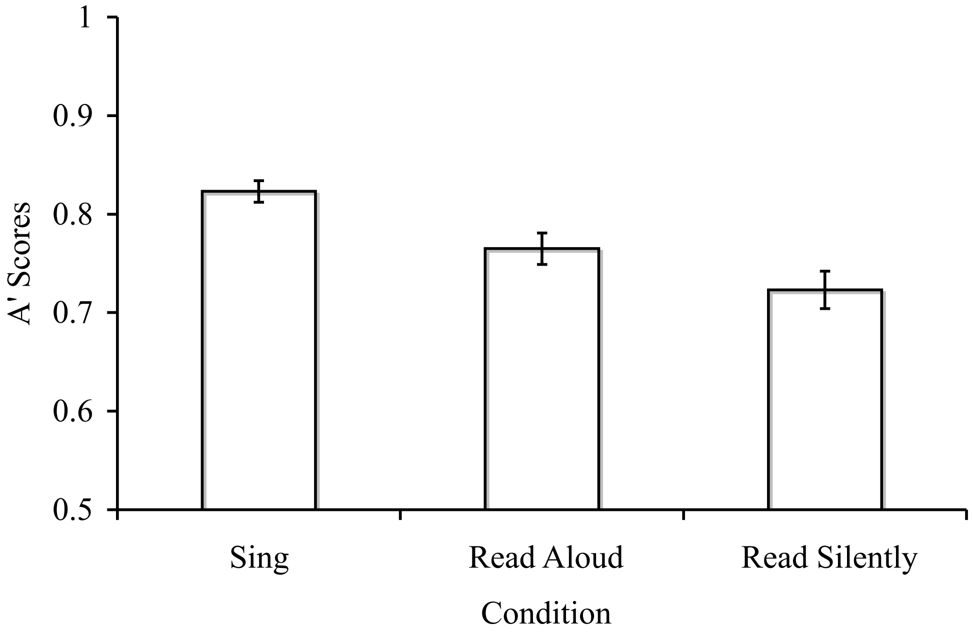


Figure 2. Experiment 2: The *A'* scores on the recognition test as a function of condition (Sing, Read Aloud, Read Silently); error bars represent one standard error.

2.4.3 Discussion

The distinctiveness account predicts better memory performance based on the number of qualitative and/or quantitative distinct elements available to distinguish studied items from one another and from unstudied items. On the grounds that singing aloud includes distinct elements in addition to articulation and audition, we reasoned that singing items at study would produce better subsequent memory performance than reading items aloud. This is exactly what we found. Using *A'* scores, we observed that while reading aloud at study increased item discriminability relative to reading silently (MacLeod et al., 2010), singing at study generated a further increase in item discriminability relative to reading aloud. In the *A'* scores, there was a production effect of .100

for items that were sung and a production effect of .046 for items that were read aloud. Thus compared to reading aloud, singing produced a further .054 increase in memory performance.[1]

2.5 EXPERIMENT 3

Experiments 1 and 2 demonstrated that other forms of production, such as reading aloud loudly and singing, result in larger subsequent memory benefits than reading aloud. This clearly establishes that there is nothing special about reading aloud *per se* that makes it a particularly good cue to relative distinctiveness; instead, relative distinctiveness depends critically on the number of qualitative and quantitative distinct elements available in the production. In Experiment 1, we argued and demonstrated that increased intensity of a verbal production serves as a cue to distinctiveness that can be used to distinguish items read aloud loudly from those read aloud in a normal voice. In Experiment 2, we argued and demonstrated that the availability of multiple potential cues to distinctiveness can be used to distinguish items sung aloud from those read aloud in a normal voice.

In Experiment 3, we employed a four level mixed-list design in which the study instructions were to Sing, Read Aloud Loudly, Read Aloud, and Read Silently. This allowed us to compare the relative effectiveness of three forms of

---

[1] Originally, we conducted a pilot study with 15 participants, which was identical to Experiment 2 with the exception that we could not calculate separate foil false alarm rates (all of items in the yes/no recognition phase were in yellow). When the mean proportions of yes responses were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor, we found the same pattern of results reported in Experiment 2: There was a significant main effect of condition, $F(2, 28)=16.436$, $MSe=.008$, $p<.001$, and all contrasts were also significant, *p's* <.034.

production (Sing, Read Aloud Loudly, Read Aloud) on memory performance. Whereas reading items aloud loudly offers increased intensity of audition relative to reading items aloud in a normal voice, singing offers the additional potential for dynamic alterations in intensity, frequency, and/or timbre (Roederer, 2008). It thus follows that when pitted against one another, singing should be an even more effective cue to relative distinctiveness than reading aloud loudly. And, of course, based on the results of Experiments 1 and 2, both singing and reading aloud loudly should be better cues to distinctiveness than reading aloud in a normal voice.

2.5.1 Method

Participants

Twenty-two undergraduate students (6 males, 16 females) participated in this experiment in exchange for credit toward their grade in an eligible Psychology class at Dalhousie University. The experiment was run in one session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language.

Stimuli and Apparatus

The stimuli and apparatus were identical to Experiments 1 and 2, except there were four rather than three instruction conditions: Sing, Read Aloud Loudly, Read Aloud, and Read Silently. Because of the additional instruction condition in this experiment, a further 80 words were selected from the Paivio, Yuille, and Madigan (www.math.yorku.ca/SCS/Online/paivio/) word generator. When these words were combined with the 240 words used in the previous experiments, they had a mean word length of 5.56 letters and 1.69 syllables. The words had a mean Kucera-Francis (1967) word frequency of 65.92, a mean imagery rating of 5.49, a mean concreteness rating of 5.38, and a mean meaningfulness rating of 6.38. Because there were four instruction conditions, the study words were printed in a red, yellow, blue, or white colored font; these represented the four instruction conditions: Sing, Read Aloud Loudly, Read Aloud, and Read Silently. Similarly, for the yes/no recognition phase, words were printed in red, yellow, blue, or white colored font.

Prior to running each participant, custom software randomly distributed the 320 words into four study lists (Sing, Read Aloud Loudly, Read Aloud, Read Silently; n=40) consisting of a total of 160 words and four foil lists (n=40) consisting of a total of 160 words.

Procedure

The general procedure was identical to Experiments 1 and 2 with the exception that there were four instruction conditions: Sing, Read Aloud Loudly, Read Aloud, and Read Silently, rather than three instruction conditions.

As in the previous experiments, before beginning the experiment, participants were given verbal instructions, which were later re-iterated on the computer monitor. Participants were told that they would be presented with a series of words printed in red, yellow, blue, or white. We used a latin-squares design to counterbalance instruction condition and color across participants; this produced four counterbalanced conditions. When explaining the four instruction conditions, the experimenter told participants to sing as they would in any other context (e.g., in the car, in the shower); read more loudly than they would in normal day-to-day interaction (e.g., so that a friend on the other side of room could hear them); read aloud in a normal voice (e.g., as if they were reading a novel aloud to themselves); and read silently such that they made no overt vocalization. The experimenter also provided examples of each of the instruction conditions.

As in the previous experiments, prior to beginning the study phase, there was a familiarization phase and a practice phase, each of which consisted of 20 trials. This was followed by the study phase, which consisted of 160 trials. There were 40 trials in each of the four conditions (Sing, Read Aloud Loudly, Read Aloud, Read Silently); these were randomly intermixed. Immediately following the study phase, there was a yes/no recognition phase. The yes/no recognition phase consisted of 320 trials: the 160 words presented in the study phase as well as 160 foil words; again, these were randomly intermixed.

2.5.2 Results

The mean proportions of hits and foil false alarms were calculated as described for Experiment 1 and are shown in Table 3.

Table 3

Means (and standard deviations) for the proportion of hits as well as foil false alarms as a function of condition (Sing, Read Aloud Loudly, Read Aloud, Read Silently).

Instruction Condition	**Hits**	**Foil False Alarms**
Sing	.629 (.144)	.233 (.170)
Read Aloud Loudly	.533 (.156)	.219 (.160)
Read Aloud	.542 (.139)	.303 (.139)
Read Silently	.397 (.132)	.202 (.131)

Sensitivity (*A'*) was calculated on a subject-by-subject basis and analyzed in a one-way ANOVA with condition (Sing, Read Aloud Loudly, Read Aloud, Read Silently) as a within-subjects factor. As seen in Figure 3, there was a significant effect of condition, $F(3, 63)=13.054$, $MSe=.004$, $p<.001$, ($\eta^2=.383$). Planned contrasts showed that the *A'* scores were significantly greater in the Sing condition ($M=.793$, $SD=.059$) than the Read Aloud Loudly condition ($M=.752$, $SD=.081$; $t(21)=2.829$, $p=.010$), the Read Aloud condition ($M=.697$, $SD=.078$; $t(21)=5.371$, $p<.001$), and the Read Silently condition ($M=.681$, $SD=.094$; $t(21)=5.351$, $p<.001$). *A'* scores were significantly greater in the Read Aloud Loudly condition than the Read Aloud condition, $t(21)=3.011$, $p=.007$, as well as the Read Silently condition, $t(21)=3.125$, $p=.005$. However, the *A'* scores were not significantly different in the Read Aloud condition than in the Read Silently condition, $t(21)=.628$, $p=.537$ (see Appendix E for additional analyses).

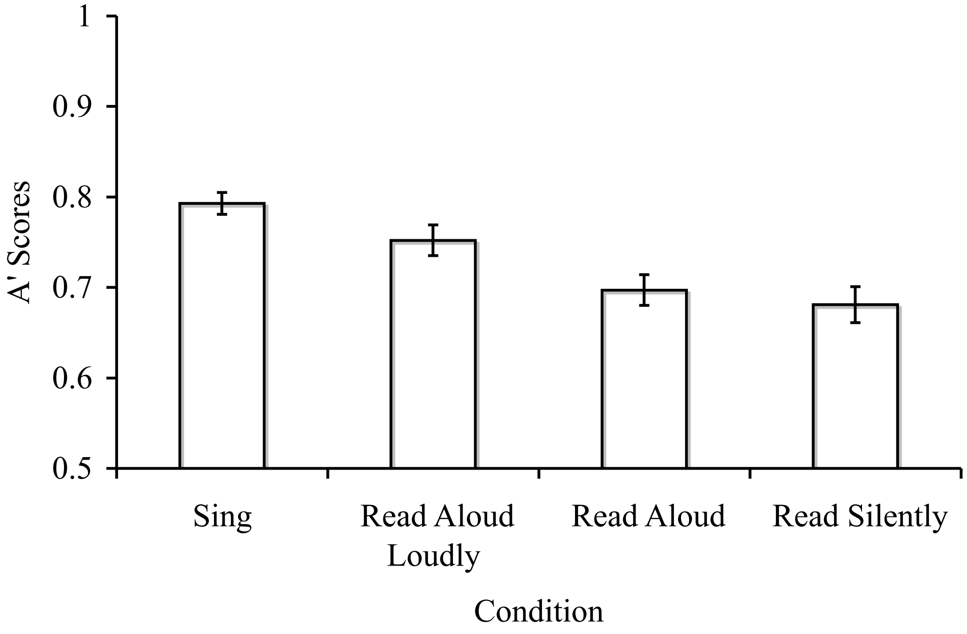


Figure 3. Experiment 3: The *A'* scores on the recognition test as a function of condition (Sing, Read Aloud Loudly, Read Aloud, Read Silently); error bars represent one standard error.

2.5.3 Discussion

In a four-level mixed design, the results of Experiment 3 demonstrate that singing produces a larger production effect than reading aloud loudly (.041) and that reading aloud loudly produces a larger production effect than reading aloud in a normal voice (.055). To our surprise, we did not find a production effect for reading aloud in a normal voice.

Despite the fact that there was no statistical difference in the *A'* scores for reading aloud versus reading silently, memory performance was in the expected direction, with higher *A'* scores in the Read Aloud condition (.697) than in the Read Silently condition (.681). In fact, there were significantly more hits for items that were read aloud (*M*=.542, *SD*=.139) compared to items that were read silently (*M*=.397, *SD*=.132; *t*(21)= 6.414, *p*<.001). There were also significantly more foil false alarms for items that were read aloud (*M*=.303, *SD*=.139) than for items that were read silently (*M*=.202, *SD*=.131; *t*(21)= 4.047, *p*=.001). Because *A'* takes into account both hit rates and foil false alarm rates, it seems likely that the difference in the foil false alarms masked the difference in hit rates for the Read Aloud and Read Silently conditions. In light of the fact that: 1) Experiments 1 and 2 replicated the standard production effect of better memory performance for reading aloud in a normal voice versus reading silently; 2) the *A'* scores were in the right direction in Experiment 3; and, 3) the production effect was significant in the hit rates in Experiment 3, we are not particularly concerned about the non-significant production effect for reading items aloud in a normal voice in the *A'* scores. And, in any case, the primary prediction of this experiment was borne out: In a mixed within-subjects design, singing at study produced greater subsequent memory than reading aloud loudly at study.

2.6 GENERAL DISCUSSION

In three experiments, we have provided corroborating evidence for claims that the production effect is related to the number of quantitative and/or qualitative distinct elements in the production that can be used as cues to relative distinctiveness (Forrin et al., 2012; MacLeod et al., 2010). Whereas Forrin and colleagues (2012) demonstrated a larger production effect for reading aloud versus other forms of verbal and non-verbal production, we demonstrated that there is nothing special about reading aloud *per se* that makes it a particularly effective cue to relative distinctiveness. Indeed, whereas reading aloud in a normal voice includes articulation and audition, other forms of vocal productions that increase the intensity of audition and/or that include other additional distinct elements, such as pitch and timbre, lead to further improvements in memory. To wit, reading items aloud loudly produced a larger production effect than reading items aloud in a normal voice (Experiment 1); singing items aloud produced a larger production effect than reading items aloud in a normal voice (Experiment 2); and, singing items aloud produced a larger production effect than reading items aloud loudly (Experiment 3).

While we believe that our data are best explained by a distinctiveness account, two alternative explanations present themselves. One possibility is that reading aloud loudly and singing are simply unusual activities and that this fact alone serves as a cue to later retrieval. This would suggest that it is not the number of distinct elements in the production that matters but how unusual the task is, with more unusual tasks resulting in better subsequent memory than more common tasks. Under this view, singing aloud might be conceived as being more unusual than reading aloud loudly, which is more unusual than reading aloud in a normal voice, which is more unusual than reading silently. While this characterization could account for our pattern of data, we do not think it is a better explanation than the relative distinctiveness hypothesis. Even though identifying distinct elements may involve some circularity of reasoning (see below), this nevertheless seems less subjective than deciding a priori which tasks might be more or less unusual. Even so, it is conceivable that the difference in the magnitude of the production effect that we observed for items sung aloud versus items read aloud might diminish or reverse if a sample of voice music students had participated instead of psychology students. Although we think that such a result is unlikely, even if it occurred, such a finding would not alter our conclusion that there is nothing special about reading aloud *per se* that improves subsequent memory performance over reading aloud silently; other vocal productions are capable of producing larger production effects.

A second alternative explanation of our findings is that reading items aloud loudly or singing items aloud at study simply takes more time and/or effort than reading items aloud in a normal voice. To the extent that the total time spent processing is correlated with subsequent memory performance (see Cooper & Pantle, 1967, for review), this would account for the pattern of data that we obtained. While this is possible, there are three reasons why this explanation is unlikely to fully account for our data. First, the presentation time of study items was not contingent on time-to-completion for the study task: Words were presented for 2000 ms in the Sing, Read Aloud Loudly, Read Aloud, and Read Silently conditions across all experiments, regardless of how much time was required to implement these tasks. Thus, the same amount of input time was available in all four conditions for participants to encode the items. Second, the processing-speed theory (Salthouse, 1996) suggests that when processing time is long, there may be *less* time for information to be encoded and elaborated. This suggests that longer output times associated with more complex productions would not necessarily translate into better memory performance. For example, if implementing the sing instruction takes relatively more time than implementing a read aloud loudly, read aloud, or read silently instruction, less time might be available to elaborate the item representation before the next study item is presented. A third argument arises from the observation that processing time cannot account for the memory difference for the Read Aloud Loudly and Read Aloud conditions. For these two conditions, participants followed the same instruction, except reading more loudly in one condition than the other; there was no difference in processing time —only a difference in the level of intensity (i.e., loudness). Arguably, reading aloud loudly might require more effort than reading aloud in a normal voice, but it is not immediately clear why this might be so. We therefore argue that any additional amount of time or effort it might take to sing versus read aloud loudly, read aloud, or read silently is unlikely to be the critical factor accounting for subsequent recognition differences in these conditions.

Instead, we favor the relative distinctiveness account forwarded by MacLeod and colleagues (Forrin et al., 2012; MacLeod et al., 2010). Even though we point to the relative distinctiveness account as the best explanation for the production effect, doing so risks circularity: Distinct processing is presumed and then confirmed by the occurrence of the result it was expected to generate (see Hunt, 2006, for review). Throughout, we have assumed that singing involves a greater number of distinct elements than reading aloud loudly, that reading aloud loudly involves a greater number of distinct items than reading aloud in a normal voice, and that reading aloud in a normal voice involves a greater number of distinct elements than reading silently. However, at no time did we manipulate or measure these distinct elements directly. We are arguing for their existence based on an intuitive understanding of the elements that are most likely to be involved in these verbal productions. For this reason, we have been particularly circumspect about identifying *which* additional elements of singing aloud might be responsible for producing a larger production effect than reading aloud loudly – there seem to be one or more potential elements that might be implicated.

One potential avenue to escape the risk of circularity in assessing the availability of cues to distinctiveness is to identify unique patterns of neural activation associated with different types of production. To this end, we can derive some support for our suppositions about the distinctiveness of singing, in particular, by noting that there are unique neural activations that distinguish the act of singing from the act of reading aloud (e.g., Epstein, Meador, Loring, Wright, Weissman, Sheppard, Lah, Puhalovich, Gaitan, & Davey, 1999; Jeffries, Fritz, & Braun, 2003; Ozdemir, Norton, & Schlaug, 2006; Stewart, Walsh, Frith, & Rothwell, 2001). Positron-emission tomography has shown that while singing and reading aloud activate overlapping brain areas, compared to reading aloud, singing produces greater activation in the right anterior temporal areas, the medial prefrontal cortex, the right superior dorsolateral prefrontal cortex, the right ventral striatum, and the limbic system (Jeffries et al., 2003). Similar findings have also been observed using functional magnetic resonance imaging. In a blocked design, Ozdemir et al. (2006) measured brain activation while participants sang or read aloud 20 bisyllabic words/phrases. While singing and reading aloud showed some overlapping regions of brain activation, compared to reading aloud, singing also showed unique activations in the mid-portions of the superior temporal gyrus (more strongly on the right than on the left) and the most inferior and middle portions of the primary sensorimotor cortex. Even though these unique patterns of activations associated with singing do not clearly identify *which* potential elements of this complex production are critical, they provide an independent means for suggesting that singing is somehow distinct from reading aloud and should therefore generate a larger production effect. In this way, brain-imaging techniques may ultimately be capable of identifying neurological underpinnings to support a distinctiveness framework.

Regardless of the success in applying brain imaging techniques to escape the potential trap of circularity, the relative distinctiveness account offers a useful framework for conceptualizing the mnemonic benefits of different productions. By specifying that the number of unique elements in a production is critical for later mnemonic success, the distinctiveness account accurately predicted that there is nothing special about reading aloud *per se* that makes it a particularly effective mnemonic strategy; indeed, there are other forms of production, such as singing and reading aloud loudly, that have even more pronounced effects on subsequent memory than reading aloud. Whether or not the distinctiveness account holds up to further scrutiny, this result on its own is important: Along with reading aloud (Ozubko et al., 2012), singing and reading aloud loudly have the potential to be powerful mnemonics with wide applicability to real-world situations. Thus, while students may benefit from reading key material aloud during study (e.g., Ozubko et al., 2012), our results demonstrate that reading aloud loudly or – better yet – singing aloud may support even larger improvements in subsequent memory. Given the relative ease with which singing and reading aloud loudly can be implemented in behavior, these strategies represent potentially useful new tools in our arsenal of mnemonic devices.

2.7 ACKNOWLEDGEMENTS

We thank Carl Helmick for designing the custom software used to randomize the items and our undergraduate participants for volunteering their time to contribute to this research. Support for this study was provided by the Natural Sciences Engineering and Research Council of Canada (NSERC) through a Vanier Scholarship to CKQ and a Discovery Grant to TLT.

CHAPTER 3: IS SINGING SIMPLY AN UNUSUAL TASK? TESTING A BIZARRENESS ACCOUNT

3.1 INTRODUCTION

In Chapter 2, we found that the production effect was greater when participants sang items compared to when they read items aloud. While we argued that this difference in the magnitude of the production effect for singing versus reading aloud was a result of distinctiveness, another possibility is that singing may be an unusual action for many participants.[2] This view suggests that it is not the number of potential distinct elements in the form of production that matters, but rather how unusual the form of production is, with more unusual forms of production resulting in better subsequent memory performance. Conceivably, singing is a more unusual form of production or action than reading aloud. Although many people read (or speak) aloud in their daily lives, many people —in fact, quite possibly the majority of people— do not sing aloud in their daily lives, especially when there is the possibility of another person hearing them.

To the extent that singing is, in fact, more unusual than reading aloud, our findings of better subsequent memory performance for items sung aloud than for items read aloud could be due to a bizarreness effect rather than a production effect. A bizarreness effect is the finding that memory performance tends to be

---

[2] I would like to thank Dr. Colin MacLeod for his feedback as a reviewer of the published manuscript included in Chapter 2, as it resulted in the decision to explore the possibility that singing may be an unusual action for many participants, resulting in better subsequent memory performance.

greater for information that is perceived as bizarre or unusual compared to information that is perceived as common (McDaniel, Anderson, Einstein, & O'Halloran, 1989; McDaniel & Einstein, 1986; McDaniel, Einstein, DeLosh, May, & Brady, 1995; for review, also see McDaniel & Bugg, 2008). There are several explanations for why bizarre or unusual items and/or tasks tend to be remembered better than common items and/or tasks. In 1989, Hirshman, Whelley, and Palij put forth the expectation violation framework. This account proposes that bizarre or unusual items violate the perceiver's expectations, resulting in a startle response that activates elaborate encoding processes and enhances orientation to general contextual cues. The expectation violation framework further suggests that the type of processing activated by unusual items results in an increased association between unusual items and general contextual cues (Hirshman et al., 1989), which is helpful at retrieval because it provides a greater potential number of available cues or information for the individual to draw upon. Following Hirshman et al. (1989), a similar explanation is that unusual items or tasks tend to produce a surprise reaction and form of arousal, which has been shown to help focus attention and cognitive control, and consequently to enhance memory performance at test (Mather & Carstensen, 2005). Similarly, Schmidt's (1991) incongruity theory proposes that unusual items trigger an attentional response which functions to orient individuals to perceiving and, subsequently, to elaborately encoding unusual items. Again, this increased attentional focus and elaborate encoding aids in enhancing memory performance at test.

A common theme of the above theoretical views is that if an item is unusual for a participant, it will likely be processed more *elaborately* than common items within the same list. This preferential elaborate processing of unusual items may be a result of a startle response, increased arousal, increased attentional focus, or a combination of the above (Hirshman et al., 1989; Mather & Carstensen, 2005; Schmidt, 1991). Regardless of the specific underlying mechanism(s) of the bizarreness effect, if items are unusual to a participant, he or she will engage in preferential processing of the items and, as a result, they will be better remembered. However, if items are no longer unusual for participants, the items should not necessarily attract preferential processing and consequently exhibit no memory advantage over common items.

At first glance, the bizarreness effect and its associated theoretical accounts may be conceptualized as being similar to a distinctiveness account; however, there are several differences that distinguish the bizarreness effect and its associated theoretical views from a distinctiveness account. Several theoretical accounts of the bizarreness effect (e.g., the expectation violation framework, incongruity theory) suggest that unusual items are remembered better than common items because unusual items generate a startle response, increased attentional focus, or increased arousal –all of which function to increase *elaborate processing or the strength of the unusual item*. In contrast, if a distinctiveness account were applied to the bizarreness effect, it would suggest that the *distinct elements* of unusual items are processed and encoded at study; a distinctiveness account does not postulate elaborative encoding as a result of increased arousal or attention. In addition, in contrast to a distinctiveness account, theoretical views associated with the bizarreness effect (i.e., the incongruity theory) do not tend to make any assumptions about possible mechanisms operating at test (i.e., the use of retrieval cues).

In a distinctiveness framework, the distinct elements of items that are encoded at study are later used as retrieval cues at test. In contrast, the bizarreness account suggests that it is the strength of items at encoding and any associated contextual cues/information that functions to enhance memory performance at test. With regard to enhanced memory performance, the distinctiveness account focuses on the interaction of processes at encoding and retrieval, whereas the theoretical accounts associated with a bizarreness effect focus only on the processes that occur at encoding. Furthermore, theoretical accounts associated with the bizarreness effect emphasize that bizarreness is defined relative to the individual, whereas distinctiveness is defined as relative to the context. Based on a distinctiveness account, in a mixed-list, within-subjects study design, items that are sung will be remembered better because they are in the context of other less distinct items (i.e., items read aloud, items read silently). In contrast, based on a bizarreness account, in a mixed-list, within-subjects study design, sung items will be remembered better only to the extent that they are perceived as bizarre by the individual participant.

3.2 EXPERIMENT 4

The goal of Experiment 4 was to determine whether the greater production effect for singing versus reading aloud can be attributed to the fact that singing may be an unusual or “bizarre” task for many people as opposed to a task that facilitates distinctiveness. We replicated the methodology of Experiment 2, but required that participants have at least one year of singing experience. Our rationale was that the act of singing should not be unusual for participants who have a history of singing experience. In a within-subjects study design, experienced singers were presented with a study phase followed by a test phase. In the study phase, words appeared one at a time in one of three coloured fonts with each colour representing a particular instruction condition: Sing, Read Aloud, and Read Silently. Immediately following the presentation of all study items, these experienced singers completed a yes/no recognition test, which included all study items as well as an equal number of foil items not presented at study.

If singing is simply an unusual task for many people, then experienced singers should be less likely than our participants in Experiment 2 to exhibit preferential processing at encoding and, consequently, a smaller or non-significant difference in memory performance for sing and read aloud items at test (Sing ≈ Read Aloud > Read Silently). In contrast, if some other mechanism, such as distinctiveness is driving the greater production effect for singing compared to reading aloud, then experienced singers should show the same pattern of memory performance as participants in Experiment 2: Sing > Read Aloud > Read Silently.

3.2.1 Method

Participants

Sixteen students (4 males, 12 females) participated in this experiment in exchange for credit toward their grade in an eligible Psychology class at Dalhousie University or for $10.00 compensation. The experiment was run in one session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language. Furthermore, all participants were required to have a minimum of one year of singing experience as a performer, member of a choir, or equivalent. The participants in the current experiment had between 2 and 20 years of singing experience with a mean of 8.19 years (*SD*= 6.27) of singing experience. Participants most commonly reported singing experience through participation in choirs. All 16 participants reported that singing was a normal, comfortable action for them that they engaged in on a daily basis.

Stimuli and Apparatus

The stimuli and apparatus were identical to Experiments 1 and 2 in Chapter 2.

Procedure

The general procedure was identical to Experiment 2 in Chapter 2, except that information related to participant's singing experience was collected via a verbal questionnaire prior to beginning the experiment. Each participant was asked the following three questions: 1) How many years of singing experience do you have? 2) What type of singing experience do you have? 3) Would you consider singing to be a comfortable, everyday task for you?

3.2.2 Results

A hit was defined as a 'yes' response to a studied word; a false alarm was defined as a 'yes' response to an unstudied foil word. The mean proportions of hits and foil false alarms are shown in Table 4.

Table 4
Means (and standard deviations) for the proportion of hits as well as foil false alarms as a function of condition (Sing, Read Aloud, Read Silently).

Instruction Condition	**Hits**	**Foil False Alarms**
Sing	.591 (.128)	.137 (.102)
Read Aloud	.484 (.122)	.121 (.084)
Read Silently	.353 (.113)	.119 (.095)

Sensitivity (*A'*) was calculated on a subject-by-subject basis and analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. As seen in Figure 4, there was a significant effect of condition, *F*(2, 30)=11.411, *MSe*=.003, *p*<.001, (η^2=.432). Planned contrasts showed that the *A'* scores were significantly greater in the Sing condition (*M*=.823, *SD*=.067) than in both the Read Aloud condition (*M*=.788, *SD*=.067; t(15)=2.720, *p*=.016), and the Read Silently condition (*M*=.736, *SD*=.063; *t*(15)=4.414, *p*=.001). The *A'* scores were also significantly greater in the Read Aloud condition than in the Read Silently condition, *t*(15)=2.450, *p*=.027 (see Appendix F for additional analyses).

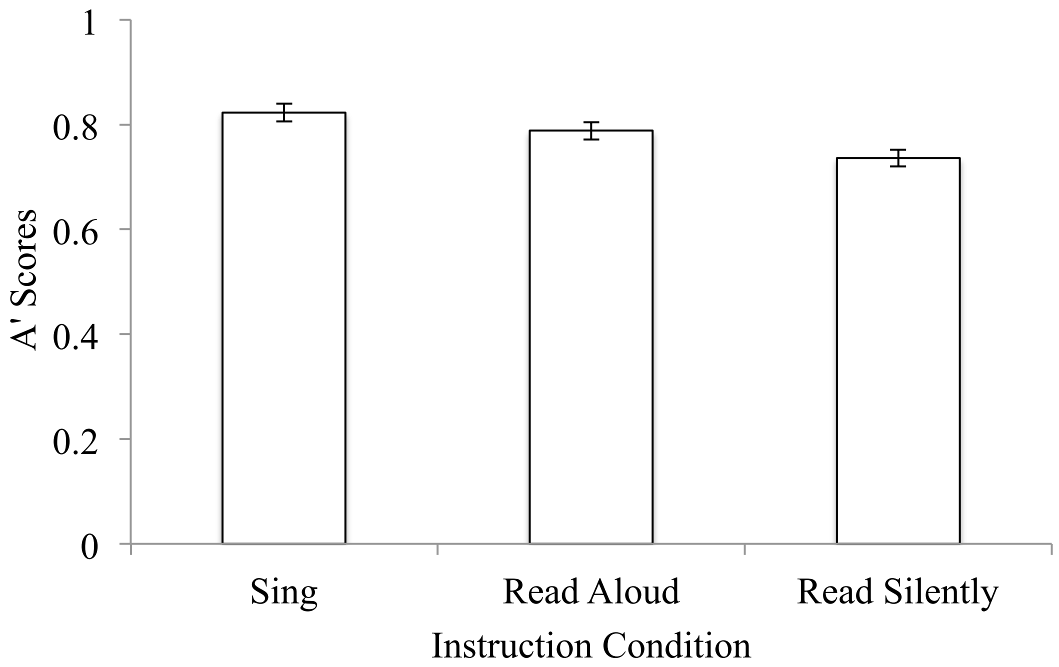


Figure 4. Experiment 4: The *A'* scores on the recognition test as a function of condition (Sing, Read Aloud, Read Silently); error bars represent one standard error.

Next, we determined whether there is evidence of a relation between years of singing experience and the magnitude of the production effect. We used years of singing experience as an estimate for how unusual or bizarre singing would be perceived by participants. To calculate the magnitude of the production effect for Sing versus Read Silently conditions, we subtracted *A'* scores in the Read Silently condition from *A'* scores in the Sing condition. A bivariate correlation analysis was conducted to examine the relation between years of singing experience and the magnitude of the production effect for Sing versus Read Silently. As shown in Figure 5, there was no significant correlation between the years of singing experience and the magnitude of the production effect for Sing versus Read Silently (r=.362, p=.273). There was also no significant correlation between the years of singing experience and memory performance in the Sing condition, (*r*=.457, *p*=.158).

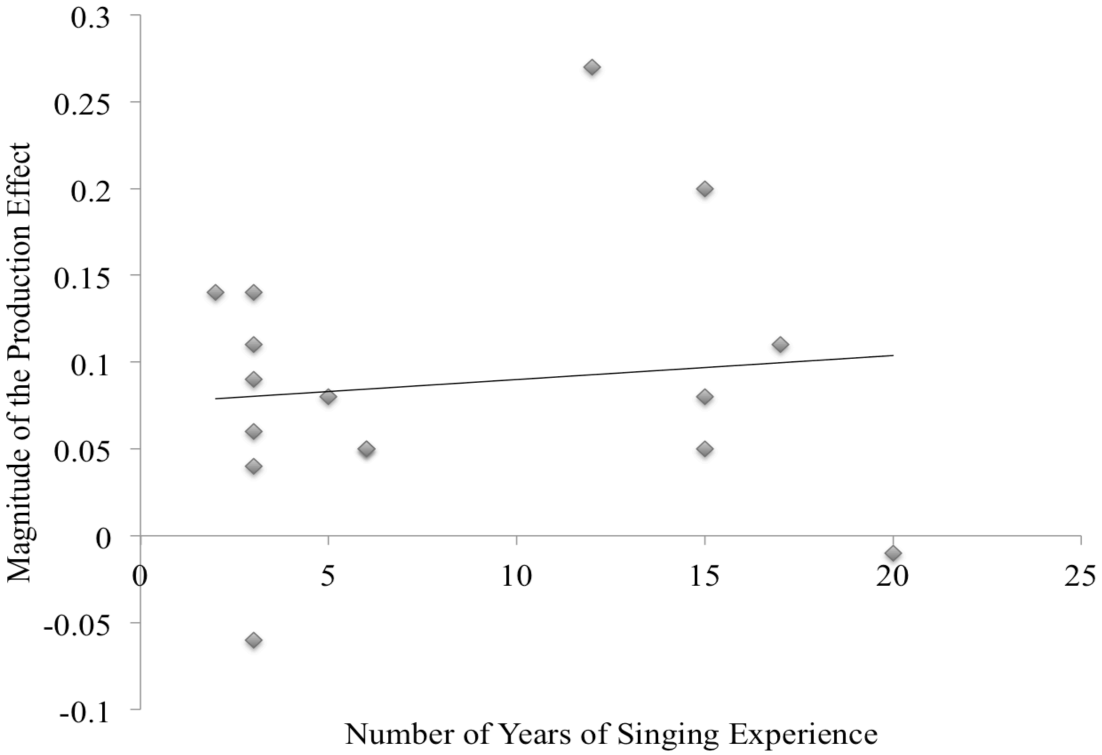


Figure 5. Experiment 4: Correlation between the magnitude of the production effect (Sing – Read Silently) and the number of years of singing experience; line represents the line of best fit.

3.2.3 Discussion

In Chapter 2, we found a greater production effect for singing compared to reading aloud. To rule out a bizarreness explanation (Hirshman et al., 1989; Mather & Carstensen, 2005; Schmidt, 1991) for our pattern of results, we replicated the methodology of Experiment 2 but required that participants have a history of singing experience to reduce any unusualness associated with singing for the individual. We predicted that if the Sing > Read Aloud > Read Silently pattern of memory performance observed in Chapter 2 was due to singing being a "bizarre" or unusual act (Hirshman et al., 1989; Mather & Carstensen, 2005; Schmidt, 1991), then the memory advantage for singing over reading aloud should be reduced or eliminated for participants who have singing experience.

In contrast, the current experiment replicated our previous findings in Chapter 2 and found a significant production effect for Sing and Read Aloud conditions with greater memory performance in the Sing condition than in the Read Aloud condition. In addition, to determine whether the pattern of results was similar to those in Chapter 2, we analyzed the *A'* scores from Experiments 2 and 4 in a mixed repeated measures ANOVA with instruction condition (Sing, Read Aloud, Read Silently) as a within-subjects factor and experiment (Experiment 2, Experiment 4) as a between-subjects factor. The main effect of experiment was not significant, $F(1, 34)=.455$, $MSe=.009$, $p=.504$, ($\eta^2=.013$) nor was the two-way interaction between instruction condition and experiment, $F(2, 68)=.459$, $MSe=.003$, $p=.634$, ($\eta^2=.013$). These findings suggest that participants from both experiments exhibited a significant production effect for both the Sing and Read Aloud conditions as well as greater memory performance for the Sing condition versus the Read Aloud condition. Furthermore, as exhibited by comparing the *A'* scores for the current experiment to Experiment 2, the magnitude of the production effect for sing versus read silent items was .09 in the current experiment versus .10 in Experiment 2, and the magnitude of the production effect for read aloud versus read silently was numerically identical in both experiments (.05). Thus, compared to reading aloud, singing produced a further .05 increase in memory performance, which is numerically identical and indistinguishable by statistical analysis. Although caution is always needed when interpreting *null* results, these findings underscore our conclusion that participants selected for singing experience produced a pattern of results similar to that reported in Chapter 2: Sing > Read Aloud > Read Silently.

From a qualitative perspective, all 16 participants in the current experiment reported that they considered singing to be a usual, everyday task that they were comfortable engaging in in their daily lives; they stated that they did not consider singing aloud to be any more unusual than reading aloud. To the extent that using participants who were comfortable with singing functioned to reduce the unusualness or bizarreness of singing compared to participants in our previous experiments, the current experiment suggests that the pattern of results obtained in Experiments 1 through 4 cannot be fully explained by a bizarreness effect and its associated theoretical views (e.g., Hirshman et al., 1989; Mather & Carstensen, 2005; Schmidt, 1991). If a bizarreness effect was driving the Sing > Read Aloud > Read Silently pattern of memory performance observed in Chapter 2 of this dissertation, we would have expected there to be a smaller or non-significant difference in memory performance for Sing versus Read Aloud conditions when using participants with a known history of singing experience. Also, to the extent that years of singing experience can estimate the level of unusualness or bizarreness of singing for individual participants, we would have also expected to observe a negative relation between the years of singing experience and the magnitude of the production effect for Sing versus Read Aloud conditions. That is, the less unusual or bizarre the task for participants (i.e., estimated by a greater number of years of singing experience) the smaller the magnitude of the production effect for Sing versus Read Aloud items.

We found no significant correlation between the years of singing experience and the magnitude of the production effect for Sing versus Read Silently (or years of singing experience and overall memory performance for sing items). Perhaps more importantly, we found that the numerical value of the correlation between the years of singing experience and the magnitude of the production effect for singing was positive rather than negative, which is contrary to the predictions of the theoretical views associated with a bizarreness account, This suggests that participants' level of singing experience is unlikely to be related to the effectiveness of singing on the production effect and therefore that our findings do not provide support for the predictions of a bizarreness account. Theoretical views of the bizarreness effect converge on the prediction that the less bizarre singing is to a participant, the smaller the subsequent memory advantage should be for words that are sung aloud versus read aloud or read silently. In contrast, a distinctiveness account does not predict such an association: The magnitude of the production effect should be related to the number of distinctive elements relative to the encoding context, regardless of the participant's history of singing experience.

That said, because we had difficulties recruiting participants with a history of singing experience, our results from the correlational analysis above are limited by a very small sample size (n=16). To find a significant association, correlational analyses typically require a much larger sample size. In addition, given that participants in the current study were university students, the range of singing experience was relatively narrow (2 to 20 years). For instance, it would be unlikely for a university student to have over 20 years of singing experience. If a future study modified the inclusion criteria of the current experiment, it would be possible to recruit a larger sample size with a wider range of singing experience. Despite these limitations associated with the correlational analysis, the mixed repeated measures ANOVA showed that there was no significant difference between the pattern of the results observed in Chapter 2 and those of the current experiment, suggesting that even experienced singers exhibit a greater production effect for singing compared to reading aloud. Moreover, despite our small sample size, in contrast to the negative correlation predicted by a bizarreness account, we found a numerically positive correlation between the years of singing experience and the production effect for singing. Thus, even if we increased sample size, the relation between years of singing experience and the production effect for singing would be unlikely to change in directionality (i.e., from being numerically positive to numerically negative).

Conclusion

The current experiment determined that experienced singers showed better subsequent memory for words sung aloud than for words read aloud or read silently. Even though their experience with singing likely made the act of singing relatively less bizarre than it might be for the general population, their pattern of results was the same as reported in Chapter 2. Although this finding does not provide direct support for the distinctiveness account that we favoured as our explanation of the Chapter 2 results, it does rule out bizarreness as a viable alternative.

We thus continue to favour a distinctiveness account of our production effects, which show greater subsequent recognition of items sung aloud than items read aloud or read silently. The distinctiveness account of the production effect does not make any assumptions regarding the degree of unusualness of the production task and the subsequent elaborate encoding or strength of items at study. Instead, this account focuses on the interaction between the distinct elements inherent in the study items and the subsequent use of these distinct elements as retrieval cues at test. As discussed in Chapter 2, singing likely involves a greater number of potential distinct elements than reading aloud: When participants sing words at study, they encode the various dynamic and item-specific elements (e.g., pitch, timbre, and intensity; Roederer, 2008) associated with each word; at test, participants use these distinct elements as retrieval cues. Compared to items read aloud and items read silently, sing items may be given preferential processing; however, based on the results of the current experiment, we suggest that this preferential processing is more likely to be the result of the encoding of distinct item-specific elements (e.g., pitch, timbre, and intensity; Roederer, 2008), than it is to be the result of bizarreness.

CHAPTER 4: DOES PRODUCTION TIME INFLUENCE THE PRODUCTION EFFECT FOR SINGING?

4.1 INTRODUCTION

As discussed in Chapter 2, it is conceivable that singing items aloud may take longer than reading items aloud or silently, and that the time spent in production (i.e., producing an item) might predict subsequent memory performance, apart from any elements of distinctiveness. To the extent that longer production time (i.e., the amount of time it takes to produce an item) translates into greater processing (see Cooper & Pantle, 1967, for review) and greater subsequent memory performance, the pattern of results obtained in the previous four experiments (Sing > Read Aloud > Read Silently) could possibly be explained by a longer production time for sing versus read aloud items. That is, the behavioural act of singing may require more time to produce an output response compared to reading aloud, resulting in greater time spent processing (Cooper & Pantle, 1967) and greater subsequent memory performance for sing items compared to read aloud items.

In fact, a recent experiment conducted in our lab (Palmer, unpublished) incorporated a key depress and release to measure the self-reported production time: On each trial, participants were required to depress a mouse key when they began to sing, read aloud, or read silently, and to release the mouse key when they were finished. Even though reading silently does not require overt production, requiring participants to depress and release a mouse key functioned to provide a baseline of output response time that could be compared to the Sing and Read Aloud conditions, which both require overt production. His recognition data replicated our previous memory performance findings: Sing > Read Aloud > Read Silently, with self-reported production durations (as measured by the mouse key depress and release) that were significantly longer in the Sing condition compared to both the Read Aloud and Read Silently conditions, which did not differ from one another. His findings show that despite a difference in memory performance for the Read Aloud and Read Silently conditions (Read Aloud > Read Silently), there is no difference in self-reported production durations (i.e., Read Aloud ≈ Read Silently), which suggests that production duration does not contribute to the production effect for items read aloud. In contrast, for items that are sung, there are differences both in memory performance (Sing > Read Aloud) and in production duration (Sing > Read Aloud), suggesting that production duration may contribute (in whole or in part) to the greater production effect for items that are sung versus items that are read aloud.

Although parsimony favours an interpretation that our pattern of findings from previous experiments (Sing > Read Aloud > Read Silently) is due to the same underlying mechanism (i.e., distinctiveness), Palmer's results raise the possibility that singing may produce greater memory performance over and above reading aloud via production duration. That is, while distinctiveness may be the sole mechanism contributing to the production effect for reading aloud, both distinctiveness and production duration may be contributing to the greater production effect for singing compared to reading aloud.

The possibility of different mechanisms underlying the production effect for singing and reading aloud is consistent with electroencephalography (EEG) findings. P300 is a neural component that is measured via EEG and considered to be indicative of distinctiveness. Hassall, Quinlan, Turk, Taylor, and Krigolson (2016) found a significantly greater P300 response for Sing and Read Aloud conditions (both required a vocalization response) compared to the Read Silently condition (did not require a vocalization response), with no significant difference in the amplitude of the P300 response for the Sing and Read Aloud conditions. Similar to Palmer, the data of Hassall and colleagues (2016) suggest that while distinctiveness may underlie the production effect for singing and reading aloud, another mechanism, such as production duration, may underlie the *greater* production effect for singing compared to reading aloud. Accordingly, if distinctiveness and a difference in production duration for sing versus read aloud items can best explain the greater memory performance for sing versus read aloud items, it follows that manipulations of production duration should influence the effect on memory of singing aloud but not of reading aloud.

4.2 EXPERIMENT 5

The goal of Experiment 5 was to determine whether the greater production effect for singing versus reading aloud can be attributed to the act of singing inherently requiring more time than reading aloud (see Palmer, unpublished). We replicated the methodology of Experiments 2 and 4 with the exception of study instructions. In the current experiment, we explicitly instructed participants to control the length of production time associated with sing and read aloud items. In contrast to previous experiments, which allowed participants to sing and read items aloud according to individual preferences, in Experiment 5, participants were instructed to sing *quickly*, read aloud *slowly*, and read silently.

We did not measure the speed of production in this experiment (this study was completed before Palmer's was performed). Nevertheless, Palmer's self-report measure of production times suggests that participants have awareness of production time durations that can be manifest in behaviour (e.g., mouse key onset/offset) – regardless of whether those production times are prompted by production instructions alone or also by expectation. To the extent that participants thus have conscious access to their subjective experience of production onset and offset, they should be able to manipulate production durations through top-down control. This control might not be sufficient to reverse the pattern of production durations that Palmer reported (i.e., he reported Sing > Read Aloud); but they should be sufficient to reduce this difference. Under this assumption, we reasoned that successful implementation of our instructions to sing quickly and read aloud slowly would reduce the memory advantage otherwise observed for singing if – and only if – production time is a critical determinant of subsequent recognition performance.

Similar to the previous four experiments, the current experiment used a within-subjects study design, wherein participants were presented with a study phase followed by a test phase. In the study phase, words appeared one at a time in one of three coloured fonts, which each represented a particular instruction condition: Sing Quickly, Read Aloud Slowly, and Read Silently. Immediately following the presentation of all study items, participants completed a yes/no recognition test, which included all study items as well as an equal number of foil items, which were not presented at study.

4.2.1 Method

Participants

There were 43 students who participated in this experiment in exchange for credit toward their grade in an eligible Psychology class at Dalhousie University. Because we used more subtle forms of production (sing *quickly* versus reading *slowly*), we decided to increase power by including a greater number of participants in the current experiment compared to the previous four experiments (this is similar to Experiment 2C in Forrin et al., 2012, which incorporated a whisper condition). The current experiment was run in one session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language.

Stimuli and Apparatus

The stimuli and apparatus were identical to Experiments 1 and 2 in Chapter 2, as well as Experiment 4 in Chapter 3.

Procedure

The general procedure was identical to Experiment 2 in Chapter 2 and Experiment 4 in Chapter 4, with the exception that the production instructions were sing *quickly*, read aloud *slowly*, and read silently. The experimenter remained in the room during the familiarization and practice phases to ensure that participants understood and followed the instructions appropriately.

4.2.2 Results

A hit was defined as a 'yes' response to studied words; a false alarm was defined as a 'yes' response to unstudied foil words. The mean proportions of hits and foil false alarms are shown in Table 5.

Table 1
Means (and standard deviations) for the proportion of hits as well as foil false alarms as a function of condition (Sing Quickly, Read Aloud Slowly, Read Silently).

Instruction Condition	**Hits**	**Foil False Alarms**
Sing Quickly	.621 (.128)	.142 (.109)
Read Aloud Slowly	.563 (.186)	.150 (.119)
Read Silently	.374 (.173)	.162 (.124)

Sensitivity (*A'*) was calculated on a subject-by-subject basis and analyzed in a one-way ANOVA with condition (Sing Quickly, Read Aloud Slowly, Read Silently) as a within-subjects factor. As seen in Figure 6, there was a significant effect of condition, $F(2, 84)=39.598$, $MSe=.004$, $p<.001$, $\eta^2=.485$. Planned contrasts showed that the *A'* scores were significantly greater in the Sing Quickly condition ($M=.833$, $SD=.061$) than in both the Read Aloud Slowly condition ($M=.802$, $SD=.088$; $t(42)=2.678$, $p=.011$), and the Read Silently condition (*M*=.710 *SD*=.092; *t*(42)=8.555, *p*<.001). The *A'* scores were also significantly greater in the Read Aloud Slowly condition than in the Read Silently condition, *t*(42)=5.481, *p*<.001 (see Appendix G for additional analyses).

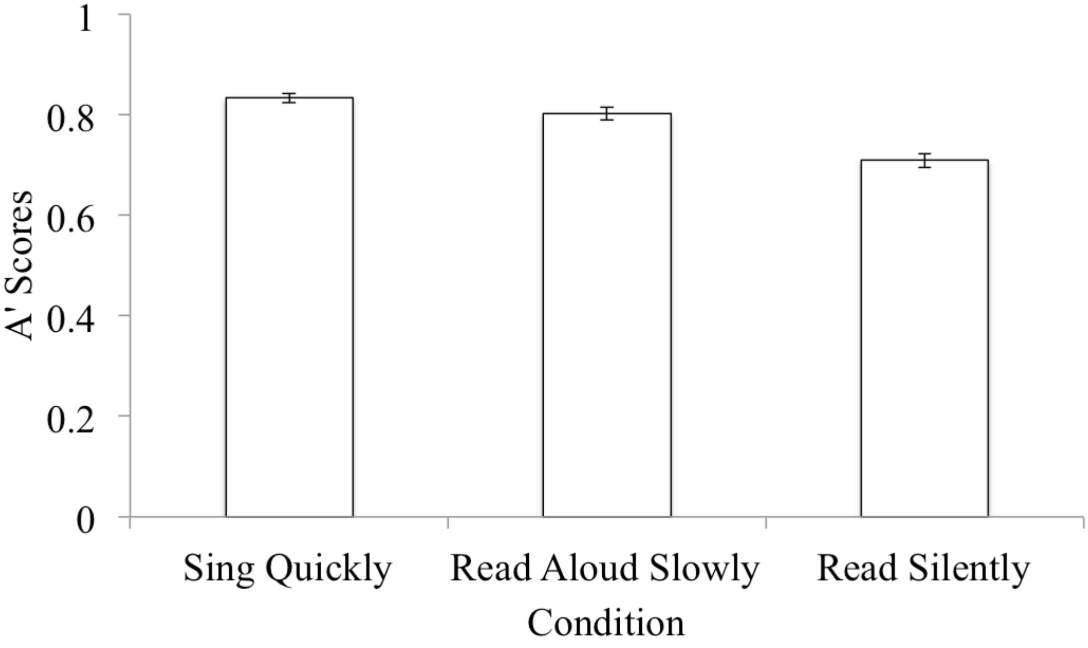


Figure 6. Experiment 5: The *A'* scores on the recognition test as a function of condition (Sing Quickly, Read Aloud Slowly, Read Silently); error bars represent one standard error.

4.2.3 Discussion

The goal of Experiment 5 was to determine whether the greater production effect for singing versus reading aloud could potentially be explained by relative output response time. Consistent with our findings in Experiments 2, 3, and 4, we found overall better subsequent memory for items that were sung aloud at study rather than read aloud. This was true despite our explicit instruction to sing *quickly* and read aloud *slowly*. This finding hints that the memory advantage for singing aloud versus reading aloud is not wholly dependent on differences in production time. To reinforce this conclusion, we compared memory performance in the current experiment (Sing Quickly, Read Aloud Slowly, Read Silently) to that of Experiment 2 (Sing, Read Aloud, Read Silently). We analyzed the *A'* scores in a mixed repeated measures ANOVA with instruction condition (Sing/Sing Quickly, Read Aloud/Read Aloud Slowly, Read Silently) as a within-subjects factor and experiment (Experiment 2, Experiment 5) as a between-subjects factor. The main effect of experiment was not significant, $F(1, 61)=.498$, $MSe=.011$, $p=.483$, ($\eta^2=.008$) nor was the two-way interaction between instruction condition and experiment, $F(2, 122)=2.242$, $MSe=.004$, $p=.111$, ($\eta^2=.035$). Despite our instruction to participants to control their speed of production, there was no significant difference between the pattern of results in the current experiment and that of Experiment 2. Taken at face value, these findings reinforce our suggestion that the pattern of results from Experiments 1 through 5 cannot be fully accounted for by a difference in the production duration for sing versus read aloud items. If production duration could explain our pattern of results, we would have expected to observe a reduction in the magnitude of the production effect for singing compared to reading aloud in the current experiment.

Our conclusion that differences in production duration are not primarily responsible for differences in memory performance for items sung aloud and items read aloud rests on the assumption that participants were able to control their production speed in accordance with instruction. We did not, however, independently verify production rates. While there would certainly be some value in collecting measures of production speed, it is not immediately obvious what the best method would be for doing so – especially to the extent that we wish to compare singing aloud to both reading aloud and reading silently. On the one hand, Palmer's method of collecting subjective measurements of production duration allows for measurement of both overt (singing aloud, reading aloud) and covert (reading silently) productions, but could be contaminated by demand characteristics. On the other hand, objective measurements cannot be readily obtained for both overt (sing, read aloud) and covert (read silently) forms of production. This is true whether a voice key were to be used (it could not be triggered for items read silently) or neural imaging techniques (which cannot easily accommodate the motor movements required for overt productions).

Admittedly, a difficulty with our interpretation arises from the fact that we did not independently verify production times. Had there been an effect of production instruction on subsequent recognition, we would have a stronger claim to the effectiveness of the instruction. As it is, the result is subject to the criticism that participants may not have followed the instructions as requested by the experimenter. This concern is underscored by the finding that reading aloud slowly did not result in greater memory performance than reading aloud (without a production speed instruction). In other words, reading aloud at a presumed slower-than-normal pace produced no improvement in subsequent memory performance. Similarly, singing quickly did not result in worse performance than singing (without a production speed instruction). In other words, singing at a faster-than-normal pace produced no reduction in subsequent memory performance.

Although we tested the view that longer production time (i.e., the amount of time it takes to produce an item) might translate into more time spent processing (see Cooper & Pantle, 1967, for review) and greater subsequent memory performance, there is an alternative view. According to Salthouse's (1996) time limited principle, if initial processing operations occur quickly, there is more time for information to be processed further (e.g., elaborated), whereas if initial processing operations occur slowly, there is less time for information to be processed further. His simultaneity principle further suggests that initial slow processing operations may be lost by the time they are needed for later processing operations, resulting in impairment of later processing operations (e.g., elaboration). Accordingly, if the act of producing the item is an early stage in item processing, there may be no reason to expect longer production times to lead to better subsequent memory. Even if singing aloud does normally take longer than reading aloud, and even if our explicit production instruction failed to alter this pattern, longer production times would be predicted to result in *poorer* processing and subsequent memory performance, not better subsequent performance.

Predictions from Salthouse's theory notwithstanding, we premised our Experiment 5 on the notion that longer production times for singing aloud might account for better memory performance compared to reading aloud. Although we found no relation between production time and subsequent memory performance, our study did not measure production times. We believe, however, that our conclusion would hold regardless. This is because Palmer subsequently provided a measurement of production times – albeit, a subjective one – yet also revealed no evidence of a relation between production duration and recognition memory performance. To demonstrate this, we calculated three bivariate correlation analyses on Palmer's data to determine whether recognition memory performance (proportion hits) could be predicted by production duration at study (measure in milliseconds) for his Sing, Read Aloud, and Read Silently conditions. As shown in Figure 7, there was no significant correlation between production duration at study and subsequent memory performance at test for the Sing (r=-.081, p=.638), Read Aloud (r=-.056, p=.747), and Read Silently conditions (r=-.254, p=.136). Thus, while production duration may be longer for sing items versus read aloud and read silently items, it is unlikely[3] that there is a significant relation between production duration and subsequent recognition memory performance within each condition.

Although the current experiment and that of Palmer have methodological limitations, our findings coupled with those of Palmer show no compelling evidence of an effect of output production duration on subsequent memory performance. Both studies do, however, underscore the fact that singing produces a greater effect on memory performance than reading aloud. There remains some ambiguity with regard to *why* this is so. The most parsimonious explanation is that singing results in a greater production effect than reading

---

[3] It is possible the sample size (n=36) in the correlational analysis was too small to detect a significant relation, limiting the ability to draw definite conclusions.

aloud because of differences in distinctive processing. However, the EEG results in Hassall et al. (2016) showed no differences in the amplitude for the P300 component (an indicator of distinct processing) elicited in response to instructions to sing versus read aloud items at study. This suggests that distinctiveness alone may not be able to account for the greater production effect for sing versus read aloud.

Contrary to our predictions, Experiment 5 tested and rejected the supposition that better memory for items sung aloud versus items read aloud might be due to longer output times. Underlying this test is an implicit assumption that longer production durations enhance memory, presumably due to increased processing (i.e., elaboration). As such, in Experiment 6, we took a different tack to address the question of whether singing aloud benefits memory over reading aloud due to a mechanism other than distinctiveness. Instead of attempting to manipulate and/or measure production duration directly, we sought evidence of differences in the strength of items sung aloud versus read aloud and read silently. If singing items aloud takes longer and this increased time allows for more processing, items sung aloud should produce better memory due to the strength of their underlying memory representations.

4.3 ACKNOWLEDGEMENTS

I would like to thank Kathy Otten for her time and effort in collecting the data for Experiment 5, and Dr. Tracy Taylor-Helmick for sharing the raw data from Ian Palmer’s undergraduate research project (supervised by Dr. Tracy Taylor-Helmick).

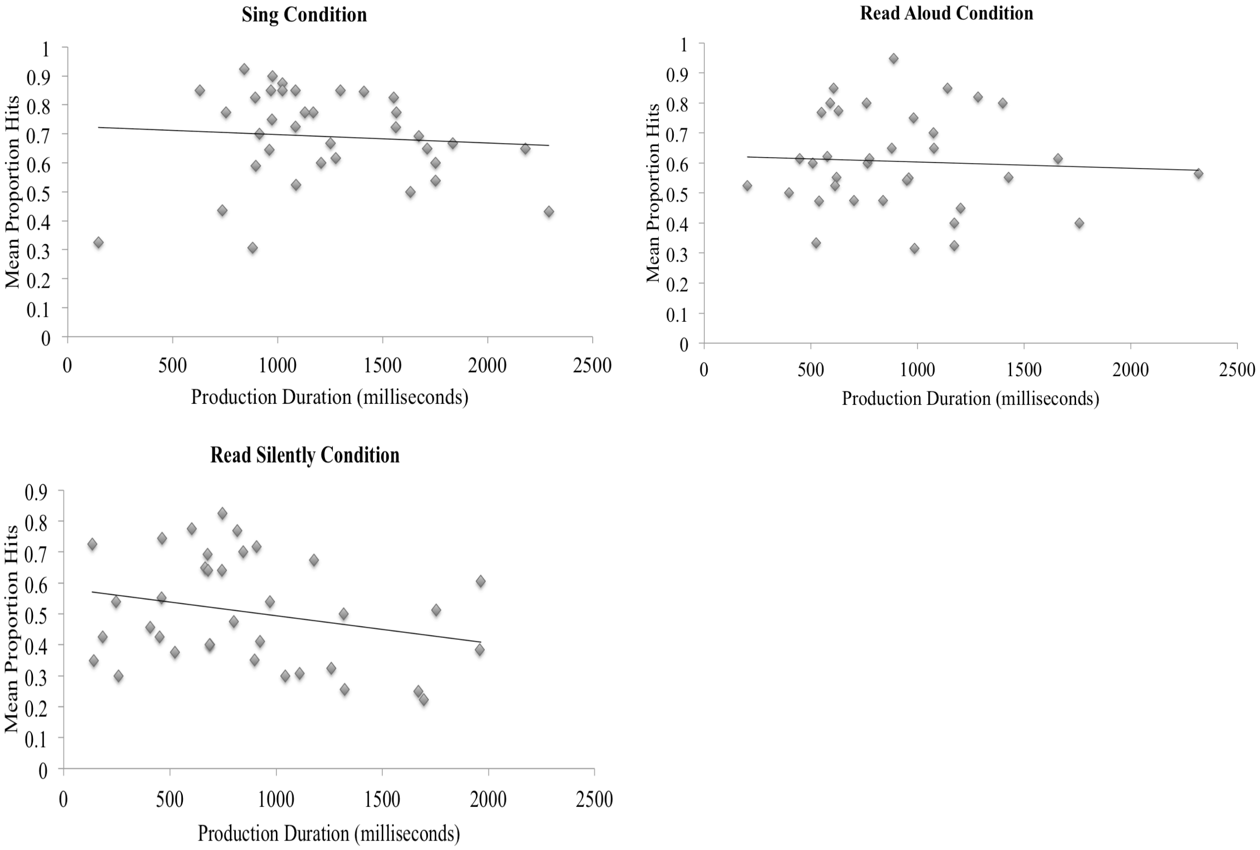


Figure 7. Palmer (unpublished): Correlations between the mean proportion hits and the production duration at study (milliseconds) as a function of condition (Sing, Read Aloud, Read Silently); lines represent the line of best fit.

CHAPTER 5: TESTING A STRENGTH-BASED ACCOUNT OF THE PRODUCTION EFFECT FOR SINGING

5.1 INTRODUCTION

Distinctiveness has been used to explain a variety of memory phenomena including the enactment effect (e.g., Engelkamp & Dehn, 2000; also see Engelkamp, 1998 for review) and the generation effect (e.g., Slamecka & Graf, 1978; see Bertsch et al., 2007, for review). Murdock (1960) stated that “the concept of distinctiveness refers to the relationship between a given stimulus and one or more comparison stimuli, and if there are no comparison stimuli the concept of distinctiveness is simply not applicable” (pp. 21). Similarly, Hunt (2006) stated, “distinctive processing is defined as the processing of difference within the context of similarity” (pp. 811). The concept of distinctiveness has also been considered from the levels of processing perspective (e.g., Lockhart, Craik, & Jacoby, 1976) such that Hunt and McDaniel (1993, pp. 423) described distinctiveness as “the processing of differences among the items of an episode.” Thus, put simply, Murdock (1960), Hunt (2006), and Hunt and McDaniel (1993) suggest that to process an item distinctively at study, the item(s) must be in the context of non-distinct items, which are used for comparison.

As described in Chapter 1, distinctiveness has typically been used to explain the production effect. Similar to the arguments above, a distinctiveness account of the production effect claims that produced items are only distinct in relation to a backdrop of non-produced items; when there is no backdrop of non-produced items, produced items are no longer distinct (MacLeod et al., 2010).

In within-subjects designs, participants are presented with both produced and non-produced items, which tend to be randomly intermixed. This type of design allows for produced items to be processed distinctively in relation to a contextual backdrop of non-produced items. In contrast, in between-subjects designs, participants are presented with either only produced or only non-produced items. Because this type of study design does not provide any contextual or relational information between produced and non-produced items, neither of the item sets should be processed distinctively (although see Jonker et al., 2013). Thus, a distinctiveness account predicts a significant production effect in within-subjects, but not in between-subjects study designs.

An alternative to a distinctiveness account of the production effect is a pure strength-based account. A pure strength-based account assumes that, compared to non-produced items, produced items are processed and encoded more elaborately at study, and thereby result in a stronger memory representation (e.g., Bodner & Taikh, 2012; Ozubko & MacLeod, 2010). Such a theoretical account is conceptually similar to a levels of processing account (Craik & Lockhart, 1972), which suggests that there is a more persistent and durable memory representation for deeply or elaborately processed items. Thus, in contrast to a distinctiveness account, a pure strength-based account would predict that produced items should be processed and encoded elaborately at study regardless of whether they are intermixed with (or in the context) of non-produced items, resulting in a significant production effect in both within- and between-subjects designs.

Many past studies have found a non-significant production effect in between-subjects designs (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010) for reading aloud versus reading silently. In the past, the absence of a significant between-subjects production effect has been interpreted as strong evidence that distinctiveness, rather than strength, underlies the production effect for items read aloud. Shortly following the collection of the data for the current experiment, Fawcett (2013) published a study wherein he found a small – albeit, significant – between-subjects production effect using a meta-analytic approach. At first glance, Fawcett's (2013) finding could be interpreted as evidence against a distinctiveness account; however, researchers have argued that the presence of a between-subjects production effect is not necessarily inconsistent with such an account.

The question at the heart of this chapter is whether the production effect for singing items is likewise attributable to distinctiveness rather than solely attributable to a pure strength based account. If so, a production effect due to singing should occur only in a within-subjects design and not in a between-subjects design. So far, the evidence presented in this dissertation converges on the notion that singing is more distinctive than both reading aloud and reading silently. We have argued that this is because singing has additional distinctive features over and above reading aloud. But it is also possible that distinctiveness derives not from the number and nature of underlying distinct features but from the verbalization itself (e.g., Hassall et al., 2016). This would suggest that singing and reading aloud are both equally distinct against the backdrop of reading silently (i.e., because both require verbalization) and that additional improvements for singing versus reading aloud are due to a separate mechanism – such as increased strength of the memory trace for items that are sung. In other words, increased recognition of items sung might be due to a combination of distinctiveness (from the verbal production) and increased trace strength.

As discussed in the previous chapter, Palmer (unpublished) found that production duration at study was significantly longer for singing compared to reading aloud and reading silently. Although production durations for reading aloud and reading silently were statistically indistinguishable, there was significantly greater memory performance for reading aloud versus reading silently (i.e., a production effect). Thus, while production duration for reading aloud may not affect processing and subsequent memory performance, the longer production duration for singing may contribute indirectly to greater elaborate encoding (via greater processing) and therefore a stronger memory trace. In Experiment 5 (Chapter 4) we manipulated production durations in an effort to undermine any tendency for greater elaboration of items sung. In Experiment 6, we take a different approach. Rather than try to influence elaborative processes, we instead attempt to measure their influence. If singing items results in stronger memory traces than reading items aloud and reading items silently, this should be evidenced in a between-subjects design. A between-subjects manipulation of production (i.e., sing, read aloud, read silently) should largely eliminate any effects of distinctiveness (see MacLeod et al., 2010) and retain only those that are due to strength. If singing improves recognition over and above reading aloud due to differences in trace memory strength, singing should continue to be a more effective production than reading aloud even when tested in a between-subjects design.

5.2 EXPERIMENT 6

The goal of Experiment 6 was to determine whether differences in elaborate encoding and strength of memory representations could account for the greater production effect for singing versus reading aloud. For Experiment 6, we used a between-subjects, rather than a within-subjects study design. Participants were assigned to one of three instruction conditions: Sing, Read Aloud, or Read Silently. In the study phase, participants viewed a series of words, one at a time. Depending upon the instruction condition of the particular participants, they were told to sing, to read aloud, or to read silently all of the study words. Immediately following the study phase, we asked participants to complete a yes/no recognition test, which included all of the study items as well as an equal number of foil items not presented at study. We used the yes/no recognition test to calculate hit rates as well as foil false alarm rates for each of the three between-subjects conditions (Sing, Read Aloud, Read Silently), which allowed us to calculate the nonparametric measure of *A'* (see Donaldson, 1996).

If singing produces a greater production effect than reading aloud due to increased processing, elaborate encoding, and strength of memory representations, we would expect to find a significant production effect for sing items compared to read aloud and read silently items. The presence of a significant production effect for sing items in a between-subjects design would suggest that distinctiveness is unlikely to be the only mechanism underlying the production effect for singing and that, in addition to distinctiveness, strength of the encoding also plays a role.

5.2.1 Method

Participants

Sixty undergraduate students participated in this experiment in exchange for credit toward their grade in an eligible Psychology class at Dalhousie University. Twenty participants were assigned to each of the three between-subjects conditions: Sing, Read Aloud, and Read Silently. The experiment was run in one session lasting approximately 30 minutes. All participants reported normal or corrected-to-normal vision and a good understanding of the English language.

Stimuli and Apparatus

As in MacLeod et al. (2010; Experiment 2), all study words were presented in white colored font and all yes/no recognition words were printed in yellow colored font. In contrast to the previous five within-subjects experiments, the current experiment used a between-subjects study design and participants were only exposed to one type of production condition (e.g., Sing or Read Aloud or Read Silently). It was therefore not necessary to use coloured fonts (i.e., red, blue, white) to differentiate the types of production conditions. Otherwise, the stimuli and apparatus were identical to Experiments 1 and 2 in Chapter 2, Experiment 4 in Chapter 3, and Experiment 5 in Chapter 4.

Procedure

Participants were assigned to one of three instruction conditions — Sing, Read Aloud, or Read Silently — so the experiment instructions varied slightly across conditions. Before beginning the experiment, participants were given verbal instructions, which were re-iterated on the computer monitor. The experimenter told participants that in the study phase they would see a series of words, one at a time. Participants in the Sing instruction condition were told that they should sing the words aloud; participants in the Read Aloud instruction condition were told that they should read the words aloud; and participants in the Read Silently condition were told that they should read the words silently (with no mouth movement or overt vocalization). The experimenter told participants that they would be required to complete a memory test following the presentation of all study words[4].

Study phase. The study phase consisted of a total of 120 trials. Each study phase trial began with a blank screen for 500 ms followed by a word in the centre for 2000 ms. Each word was selected randomly without replacement from the study list. The total duration of each study trial was 2500 ms.

Recognition phase. Upon completing the study phase, participants began the recognition phase. The recognition phase consisted of a self-paced yes/no recognition test. At the beginning of the recognition phase, instructions were

---

[4] Because participants were only exposed to one production condition (i.e., Sing, Read Aloud, Read Silently) in the current experiment, we did not think it was necessary to include a familiarization or practice phase. In the previous five within-subjects experiments, there was a familiarization phase and a practice phase, which were designed to help participants remember and internalize the colour/production condition associations (e.g., red represents a sing production instruction; blue represents a read aloud production instruction; and white represents a read silently production instruction).

presented at the top of the computer screen. Participants were instructed to press the ‘y’ key if they recognized the word from the study trials and to press the ‘n’ key if they did not recognize the word from the study trials (i.e., a foil word). All responses could be self-corrected using the backspace key and were submitted by pressing the space bar. The next recognition trial began after each response was submitted. On each recognition trial, a word was presented at the centre of the computer monitor until a response was made. In total, there were 240 recognition trials: The 120 words presented in the study phase and 120 foil words, which were intermixed randomly.

5.2.2 Results

A hit was defined as a ‘yes’ response to studied words; a false alarm was defined as a ‘yes’ response to unstudied foil words. The mean proportions of hits and foil false alarms are shown in Table 6.

Table 6
Means (and standard deviations) for the proportion of hits as well as foil false alarms as a function of condition (Sing, Read Aloud, Read Silently).

Condition	**Hits**	**Foil False Alarms**
Sing	.617 (.160)	.192 (.144)
Read Aloud	.581 (.115)	.124 (.079)
Read Silently	.608 (.146)	.204 (.143)

Sensitivity (*A’*) was calculated on a subject-by-subject basis and analyzed in a one-way ANOVA with instruction condition (Read Aloud Loudly, Read Aloud, Read Silently) as a between-subjects factor. Lower *A’* values represent lower sensitivity and higher *A'* values represent greater sensitivity (a value of .50 represents chance performance). As can be seen in Figure 8, there was no significant effect of instruction condition, *F*(2, 57)=1.267, *MSe*=.004, *p*=.290. Indeed, contrasts showed that the *A'* scores were not significantly different for the Sing (*M*=.808, *SD*=.071) and Read Aloud conditions (*M*=.828, *SD*=.047; *t*(38)= 1.030, *p*=.310) or the Sing and Read Silently conditions (*M*=.796, *SD*=.071; *t*(38)= .546, *p*=.588); nor was there a significant difference between the Read Aloud and Read Silently conditions, *t*(38)= 1.683, *p*=.101 (see Appendix H for additional analyses).

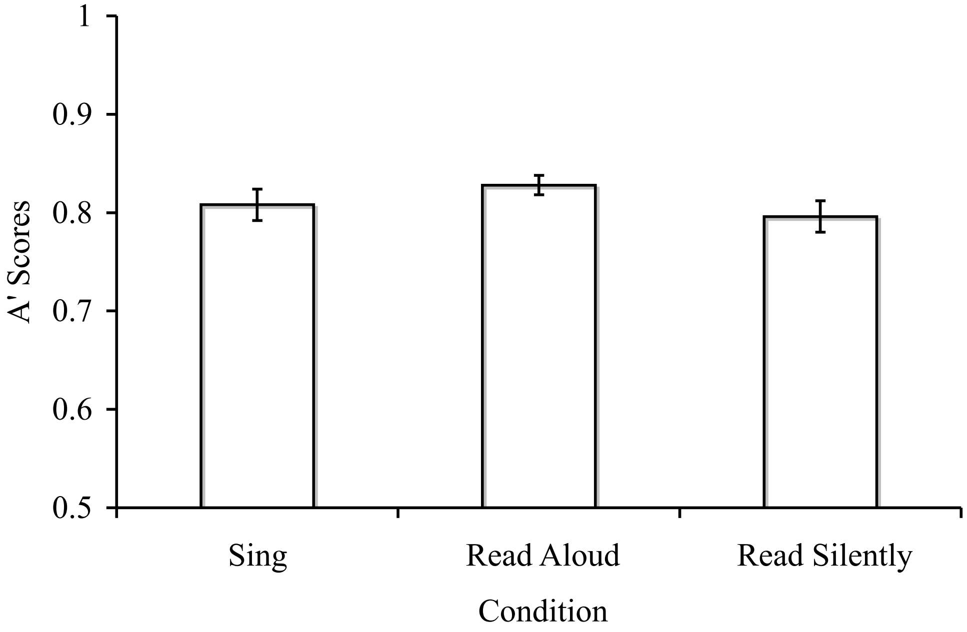


Figure 8. Experiment 6: The *A'* scores on the recognition test as a function of between-subjects condition (Sing, Read Aloud, Read Silently); error bars represent one standard error.

5.2.3 Discussion

Experiment 6 used a between-subjects study design to determine whether differences in elaborate encoding and strength of memory representations could account for the greater production effect for singing versus reading aloud. Based on previous findings using a between-subjects design (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010), we predicted that if singing induces a greater production effect than reading aloud due to differences in trace memory strength, we would find a significant between-subjects production effect for items that are sung compared to items that are read aloud and read silently. The presence of a between-subjects production effect for singing would suggests that, in addition to distinctiveness, differences in elaborate encoding and strength of memory trace also contribute to the greater production effect for singing versus reading aloud (e.g., as afforded, perhaps, by longer production durations). Using *A'* scores, we observed that item discriminability was statistically equivalent across the three between-subjects conditions indicating that there was no significant production effect for the Sing or Read Aloud conditions. These findings are in contrast to our consistent findings using a within-subjects design (see Chapters 2, 3, and 4): In a within-subjects design, production effects occur for both the Sing and Read Aloud conditions, with a larger effect for the Sing condition.

Our findings for the Read Aloud condition replicate the existing literature to demonstrate that the production effect (for reading aloud versus silently) tends to be limited by a between-subjects design (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010; although see Fawcett, 2013) consistent with a distinctiveness account (see MacLeod et al., 2010). Extending previous findings, the results of the present experiment demonstrate that the larger production effect for singing versus reading aloud is likewise limited by a between-subjects design.

This strongly suggests that singing results in better subsequent recognition than reading aloud for the same reason that reading aloud results in better subsequent recognition than reading silently: distinctiveness. In Chapter 4, we attempted to influence elaborative processes by manipulating production duration and, similar to the current experiment, we found no significant effect of production duration on subsequent memory performance. The results of the current experiment and those of Experiment 5 thus converge on a similar outcome: the greater production effect for singing versus reading aloud is more likely due to distinctiveness than to additional processing.

A distinctiveness account of the production effect emphasizes the relation between the distinct elements at study and the subsequent use of these distinct items as retrieval cues at test. As indicated in previous chapters, we suggest that singing likely involves a greater number of potential distinct elements than reading aloud. Singing and reading aloud involve articulation and audition; however, in addition to these two-shared distinct elements, we suggest that singing also involves various dynamic and item-specific distinct elements (e.g., pitch, timbre, and intensity; Roederer, 2008). Thus, at test, participants have a greater number of distinct elements to use as retrieval cues for singing compared to reading aloud.

Of course, although our experiment had a similar (Hopkins & Edwards, 1972) or larger (Gathercole & Conway, 1988; MacLeod et al., 2010) sample size compared to many previous between-subjects production studies, given the results of the meta-analysis conducted by Fawcett (2013), it is conceivable that we did not have sufficient power to detect a significant between-subjects production effect for the Sing or Read Aloud conditions. It is also possible that regardless of our sample size, we would not have found a significant between-subjects production effect. The Sing > Read Aloud > Read Silently pattern of memory performance observed in our within-subjects experiments was not numerically evident in Experiment 6, which suggests that even with additional participants, we may not have observed a significant between-subjects production effect. Alternatively, it is also possible that the small —albeit significant – between-subjects production effect found by Fawcett (2013) may not reflect a true effect. The findings of studies using a meta-analytic approach have the potential to be substantially impacted by errors including publication bias (i.e., studies showing an effect are more likely to be published than those that show no effect) and the inclusion of a small number of heterogeneous studies. At this point, there is not compelling evidence to make strong claims about the occurrence of a between-subjects effect, and it may be more valuable to interpret on a study-by-study basis, depending upon the research question and the experimental methodology.

In summary, the results from Experiment 6 converge on the notion that distinctiveness is the sole mechanism that gives rise to production effects for singing items. There is no evidence that singing items strengthens item traces relative to reading items aloud. If there were effects of singing on pure memory strength, the increased strength would have persisted in a between-subjects design. Instead, recognition was equivalent for items sung, read aloud, and read silently in our between-subjects manipulation.

CHAPTER 6: GENERAL DISCUSSION

6.1 SUMMARY OF FINDINGS

The purpose of this dissertation was to further delineate the production effect. This was accomplished by: 1) determining whether alternate forms of vocal production, such as reading aloud loudly and singing, have a greater impact on the production effect than reading aloud in a normal voice (Chapter 2), and 2) investigating the possible mechanisms underlying the influence of alternate forms of vocal production on the production effect (Chapters 3-5).

A summary of the study design, production conditions, and magnitude of the production effect relative to a Read Silently condition, is presented in Table 7 as a function of experiment. Experiment 1 included two forms of production: Reading aloud in a loud voice and reading aloud in a normal voice, and compared these two forms of production to reading silently. Conceivably, compared to reading aloud in a normal voice, reading aloud in a loud voice involves an additional potential distinct element at encoding (i.e., intensity) and thus, based on a distinctiveness account of the production effect, reading aloud loudly should result in a greater production effect than reading aloud in a normal voice. Indeed, while we found a production effect for both reading aloud loudly and reading aloud in a normal voice, the effect was greater in the Read Aloud Loudly condition compared to the Read Aloud condition.

Table 7
Summary of study design, production conditions, and magnitude of the production effect (relative to a Read Silently condition) as a function of experiment.

Production Effect						
	Study Design	Sing	Sing Quickly	Read Aloud Loudly	Read Aloud	Read Aloud Slowly
Experiment 1	Within			**.096**	**.044**	
Experiment 2	Within	**.100**			**.046**	
Experiment 3	Within	**.112**		**.071**	.016	
Experiment 4	Within	**.087**			**.052**	
Experiment 5	Within		**.123**			**.092**
Experiment 6	Between	.012			.031	

Note: Significant production effects (p <.05) are denoted in bold font.

Similar to Experiment 1, Experiment 2 included two types of vocal production, singing and reading aloud in a normal voice, and compared these two forms of production to reading silently. Compared to reading aloud, singing can be regarded as involving several additional distinct elements at encoding including intensity, frequency, and/or timbre (Roederer, 2008) and thus, based on a distinctiveness account of the production effect, singing should produce a larger production effect than reading aloud. The findings of Experiment 1 were extended such that while there was a production effect for both singing and reading aloud, the effect was larger for the Sing condition compared to the Read Aloud condition (i.e., with even better subsequent memory for items that were sung).

Experiment 3 included all three forms of production: singing, reading aloud loudly, and reading aloud in a normal voice, all of which were compared to reading silently. The goal of Experiment 3 was to determine the individual contributions of singing and reading aloud loudly to the production effect. Given that singing is conceived to consist of multiple dynamic distinct elements (intensity, frequency, and/or timbre) compared to reading aloud, and reading aloud loudly is conceived to consist of a single additional distinct element (intensity) compared to reading aloud in a normal voice, we predicted that the production effect would be greater for sing items compared to read aloud loudly items. Indeed, while both singing and reading aloud loudly produced a significant production effect, the production effect was greater for sing items compared to read aloud loudly items (Sing > Read Aloud Loudly > Read Aloud > Read Silently).

At first glance, the findings from Experiments 1-3 appeared to provide support for the distinctiveness account of the production effect; however, Experiment 4 sought to rule out an alternative explanation for our findings. That is, singing may be an unusual task for many individuals and involve similar mechanisms proposed by theoretical accounts associated with the bizarreness effect (e.g., Hirshman et al., 1989; Mather & Carstensen, 2005; Schmidt, 1991). To determine whether the unusual act of singing was driving the greater production effect for singing versus reading aloud, Experiment 4 replicated the methodology of Experiment 2 using participants who were experienced singers. We hypothesized that if the greater production effect for singing versus reading aloud is because singing is a bizarre or unusual act for many people, this difference should be smaller or non-significant when singing is no longer considered to be bizarre or unusual (Sing ≅ Read Aloud > Read Silently). Consistent with our previous findings, we found that singing items produced a greater production effect compared to reading items aloud, even in experienced singers; the magnitude of the Sing > Read Aloud difference did not differ statistically from our previous findings in Chapter 2. In addition, there was no significant association between the years of singing experience and the magnitude of the production effect for Sing versus Read Aloud conditions. Together, the results from Experiment 4 suggest that our findings cannot be adequately explained by a bizarreness effect.

The purpose of Experiment 5 was to ascertain whether the greater production effect for singing versus reading aloud could be attributed to the fact that the act of singing may inherently require a longer production duration than reading aloud (see Palmer, unpublished). Experiment 5 replicated the methodology of Experiment 2, but explicitly instructed participants to sing *quickly*, read aloud *slowly,* and read silently; this change in production instructions allowed us to test whether a difference in production duration for singing versus reading aloud could explain the results in our previous experiments. If a longer production duration for sing items is driving the greater production effect for singing, then explicitly instructing participants to reduce the production duration for sing items and increase the production duration for read aloud items should function to reduce or eliminate the greater production effect for sing versus read aloud. Experiment 5 replicated our previous finding: Singing produced a greater production effect than reading aloud with no significant difference in memory performance in Experiment 5 compared to Experiment 2 (e.g., explicit instructions with regard to production duration versus no explicit instructions with regard production duration). While there may be an inherent difference in production duration for singing versus reading aloud at study (Palmer), to the extent that our study instructions were heeded, our findings suggest that a difference in production duration for sing versus read aloud items cannot fully explain our pattern of results (i.e., Sing > Read Aloud > Read Silently). In Experiment 5, we assumed that requiring participants to control their speed of production would have revealed any effect of production duration on memory performance. However, underlying this assumption is that longer production durations result in greater memory performance via increased processing and elaborate encoding, which then leads to greater memory performance.

As such, the goal of Experiment 6 was to determine whether differences in elaborate encoding and strength of memory representations could account for the greater production effect for singing versus reading aloud. Experiment 6 replicated the methodology of Experiment 2 using a between- subjects as opposed to a within-subjects design. If singing produces a greater production effect than reading aloud due to increased processing, elaborate encoding, and strength of memory representations (perhaps via longer production durations), we would expect to find a significant production effect for sing items compared to read aloud and read silently items, even in a between-subjects design. Contrary to this prediction, using a between-subjects design in Experiment 6, we found no significant production effect for sing or read silently items (i.e., statistically, Sing = Read Aloud = Read Silently). These findings are consistent with those of the production effect for reading aloud, suggesting that the production effect is ordinarily eliminated in between-subjects designs (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010; although see Fawcett, 2013, for review), and that a common mechanism – distinctiveness – underlies the production effect for both singing and reading aloud.

6.2 CONNECTION TO A DISTINCTIVENESS ACCOUNT AND PAST FINDINGS

As discussed throughout the current dissertation, a distinctiveness account has typically been used to explain the production effect. Such an account assumes that producing an item results in a relatively more distinct memory trace, making produced items easier to retrieve at test compared to non-produced items (see Schmidt, 1991, and more recently, Hunt, 2006, for a review of distinctiveness). MacLeod and colleagues (2010) suggested that produced items have at least one distinct element that is encoded at study compared to non-produced items. They further suggested that these distinct elements function as retrieval cues to guide memory performance at test. For instance, compared to reading silently, reading aloud consists of two additional distinct elements: articulation and audition, which are encoded at study. At test, participants can use these two distinct elements to decide whether an item was studied: If participants remember saying the word aloud and/or hearing themselves say the word aloud, they can use that information to decide that the item was presented at study. Similar to reading aloud, other forms of production, such as spelling, typing, writing, and mouthing (see Forrin et al., 2012) have been suggested to involve an additional distinct motor element that is not present while reading silently. For instance, typing involves the movement of the hands depressing keys on the keyboard, whereas mouthing involves the movement of the lips forming the words without saying them aloud.

Past studies have found evidence of a graded pattern of memory performance for various forms of production, which varies according to the number of presumed distinct elements. Forrin et al. (2012) found that reading aloud at study produced greater memory performance than writing, mouthing, and whispering – all of which produced better memory than reading silently. They argued that the [Read Aloud > Writing, Mouthing, Whispering > Read Silently] pattern of memory performance could best be explained by differences in the number of distinct elements presumed to underlie each type of production. For instance, reading aloud involves two distinct elements, audition and articulation, whereas typing and mouthing only involve one distinct element: movement with the hands or mouth. Although whispering involves the same two distinct elements as reading aloud (audition and articulation), Forrin et al. (2012) suggested that whispering produced memory performance intermediate to reading aloud and reading silently because the intensity associated with whispering is weaker than the intensity associated with reading aloud in a normal voice (also see Castel, 2009, and Castel et al., 2013). Thus, distinct elements can be conceptualized in terms of quantity (i.e., the number of distinct elements; motor, motor + audition) or quality (i.e., the richness or value of the distinct element; intensity of audition), with the effect on subsequent memory depending on the total number *and* quality of distinct cues available at retrieval.

Consistent with past research (Castel, 2009, Castel et al., 2013; Forrin et al., 2012), across our five within-subjects experiments, we found that the number and quality of distinct elements available at encoding was associated with subsequent memory performance at test and consequently the magnitude of the production effect (see Table 7 for overview). Similar to reading aloud in a normal voice, singing aloud consists of articulation and audition, but in addition, it also consists of several dynamic elements including intensity, pitch, and/or timbre (Roederer, 2008). The availability and utility of the additional distinct elements associated with singing were evident in the pattern of memory performance and magnitude of the production effect across our experiments: Singing in a normal voice (regardless of speed; at a normal speed or quickly) consistently produced a greater production effect than any other form of production in within-subjects study designs. Reading aloud loudly and reading aloud slowly consistently produced a production effect that was intermediate in magnitude to singing and reading aloud in a normal voice. While reading aloud loudly or slowly does not consist of a greater number of distinct elements compared to reading aloud in a normal voice/speed, both forms of production consist of a greater quality of distinct elements compared to reading aloud in a normal voice/speed. The intensity associated with reading aloud loudly is greater than reading aloud in a normal voice (increased intensity of audition), whereas the rate of output and subsequent processing associated with reading aloud slowly are hypothesized to be greater than reading aloud in a normal voice (Cooper & Pantle, 1967). Together, our pattern of findings, coupled with previous results (Castel, 2009; Castel et al., 2013; Forrin et al., 2012), emphasizes the importance of the role of distinctiveness in the production effect. To the extent that the quantity and quality of distinct elements encoded at study function as retrieval cues at test, when there are more distinct elements encoded for a given item at study, there is a greater likelihood of recognition of that item at test.

Equally important, we found that the graded pattern of memory performance (Sing > Read Aloud > Read Silently) only occurs in within-subjects designs, suggesting that the relative comparison of items plays an important role in the production effect. Given that Hunt (2006, p.811; as well as other theorists, Hunt & McDaniel, 1993; Murdock, 1960) suggested that distinctiveness is “the processing of difference within the context of similarity,” it is not surprising that when there is no longer a backdrop of non-produced (or less distinct) items available for comparison purposes, there is no longer a significant production effect. As Fawcett (2013) pointed out, the degree to which a distinctive heuristic may be applied depends upon both the diagnostic utility of the available retrieval cues and the likelihood of their application. Between-subjects study designs reduce the probability of a production-based distinctiveness heuristic being used at test, based on our findings, this seems to be true regardless of the quantity and/or quality of distinct elements. Given that participants are exposed to both produced and non-produced items in a within-subjects study design, participants are able to use distinct retrieval cues (i.e., articulation, audition, intensity, pitch, timbre) at test; however, given that participants are only exposed to produced or non-produced items in a between-subjects study design, participants are not able to use distinct retrieval cues as an indicator of whether an item was studied.

Although it is rare (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010), researchers occasionally find a significant between-subjects production effect (for review, see Fawcett, 2013) for items read aloud versus read silently. Fawcett (2013) suggests that distinctiveness may play some role in both designs, but that participants may simply be less likely to use a distinctiveness heuristic when there is no backdrop of non-produced items (as in a between-subjects design). Given the graded pattern of memory performance across all of our experiments (Sing > Read Aloud > Read Silently), we would argue that *both* read aloud and read silently items form the backdrop against which sung items are rendered distinct. If so, this makes the distinctiveness heuristic even more unhelpful when this backdrop is missing in a between-subjects design. Our graded pattern of memory performance only occurred in within-subjects study designs, consistent with the important role that relative distinctiveness is known to play in the production effect.

6.3 ALTERNATIVE THEORETICAL VIEWS: RULED OUT

Although our results converge on the conclusion that additional distinct elements are responsible for the fact that the production effect is even larger for singing than for reading aloud, it was important to rule out alternative explanations. The literature has established that the production effect for reading aloud versus silently is due to relative distinctiveness of the production. It was incumbent upon us to likewise determine whether the especially large production effect for singing was also attributable to distinctiveness alone. In the current dissertation, we questioned whether other mechanisms such as level of experience/bizarreness, production duration/greater processing, and strength of memory representation could be contributing to the greater production effect for singing versus read aloud. Through the series of experiments reported, we were able to rule out these alternative explanations. As discussed below, findings in the wider production effect literature are also inconsistent with these alternative accounts of the production effect.

Our experiment using experienced singers (Experiment 4) replicated the pattern of findings in our previous experiments: Sing > Read Aloud > Read Silently, with no significant difference in the pattern of results across experiments. To the extent that singing experience is an indicator of level of experience or "bizarreness," we found no relation between years of singing experience and the magnitude of the production effect for singing. Our findings provided evidence against a bizarreness explanation (Hirshman et al., 1989; Mather & Carstensen, 2005; Schmidt, 1991) for the greater production effect for sing versus read aloud. The findings of Forrin et al. (2012) also are inconsistent with a bizarreness explanation of the production effect, which would predict that more unusual or bizarre forms of production would be associated with greater subsequent memory performance. Humans are constantly speaking and engaging in conversations with others. Thus, it could be argued that writing, whispering, and mouthing are inherently *more* unusual forms of production than reading aloud (i.e., many people engage in those tasks less frequently than speaking aloud). However, in contrast to the predictions of a bizarreness account, Forrin and colleagues (2012) found a smaller production effect for writing, whispering, and mouthing compared to reading aloud. Together, our findings and those of Forrin et al. (2012) provide no support for the role of a bizarreness explanation of the production effect.

We also ruled out the hypothesis that singing improves recognition over and above reading aloud simply because singing takes longer to perform (e.g., Palmer). In Experiment 5, we required participants to control their speed of production – read aloud *slowly* and sing *quickly*. We found no evidence that production duration was directly associated with subsequent memory performance. While Forrin et al. (2012) did not directly measure production duration, their results can nevertheless also speak to the relation between production duration and subsequent memory performance. Consider the amount of time it takes to read aloud versus write. It is conceivable that most (if not all) individuals would take measurably longer to write a word on paper than to read that same word aloud. Thus, to the extent that longer production durations are responsible for greater subsequent memory performance, writing should result in better subsequent memory performance than reading aloud; this was not the case. Forrin et al. (2012) found Read Aloud > Write > Read Silently, providing indirect evidence against the role of production duration in the production effect. In the same vein, reading aloud should take approximately the same amount of time, regardless of intensity. That is, reading aloud loudly, reading aloud in a normal voice, and whispering would not necessarily be expected to differ in their production durations. Our pattern of memory performance coupled with that of Forrin et al. (2012) shows: Reading Aloud Loudly > Read Aloud > Whispering > Read Silently, which is inconsistent with the predictions based on the assumed relation between production duration and subsequent memory performance. Thus, similar to our results in Experiment 5, the findings of Forrin et al. (2012) also argue against the role of production duration in the production effect.

In Experiment 6, the use of a between-subjects design revealed no significant production effect. This ruled out the possibility that singing produces an advantage over reading aloud due to stronger representations for items sung (e.g., due to increased processing associated with longer production durations). While this finding was consistent with many past studies that used a between-subjects study design to examine the production effect (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010), it was inconsistent with Fawcett (2013), who reported a numerically small but significant between-subjects production effect based on a meta-analysis. The finding of a significant between-subject effect resulted in researchers discussing alternative explanations and accounts for the production effect (for review, see Obzuko et al., 2012, as well as Bodner & Taikh, 2012, and Taikh & Bodner, 2016). While the distinctiveness account may not predict a between-subjects production effect, we think that such an effect is not entirely incompatible with a distinctiveness account; participants may simply be *less* likely to use the distinct information/elements as retrieval cues at test because there is no longer a backdrop of non-produced items. This view assumes that the likelihood of participants using a distinctiveness heuristic is not black and white and lies more along a continuum depending upon the particular study conditions. A key finding that supports the notion of the *likelihood* of participants using a distinctiveness heuristic is the larger effect size for within- compared to between-subjects study designs (Quinlan & Fawcett, unpublished). A greater production effect for within- versus between-subjects designs suggest that producing items may be distinct at study for both within- and between-subjects designs, but the extent to which that distinct information is used, as a retrieval cue at test is greater for within-subjects designs compared to between-subjects designs because there is a backdrop of non-produced items in within-subjects designs that serves as a comparison.

Further evidence against strength-based account of the production effect comes from neuroimaging studies. If the greater production effect for singing versus reading aloud was due to differences in elaborate encoding and strength of memory representation, we would expect to see neural activation in the same brain regions, but to a greater extent for sing versus read aloud. Instead, researchers have found that, compared to reading aloud, singing produces neural activations in several other brain regions than reading aloud (Jeffries et al., 2003; Ozdemir et al., 2006). Thus, our findings coupled with previous research suggest that at present, there is no strong evidence that a mechanism other than distinctiveness contributes to the production effect either for reading aloud or for singing.

6.4 COMPONENT PROCESSES UNDERLYING THE PRODUCTION EFFECT FOR SINGING

Despite the fact that our findings support a distinctiveness account of the production effect, past findings examining the component processes underlying the production effect have been considered problematic for a distinctiveness account. We now consider these component processes as they relate to our findings and the mechanisms underlying the production effect.

The magnitude (i.e., size) of the production effect is measured by subtracting memory performance for non-produced items from produced items. Two component processes can therefore impact the size of the production effect: 1) a decrease in memory for non-produced items (i.e., costs) and/or 2) an increase in memory for produced items (i.e., benefits). It is often assumed that the production effect reflects a benefit for produced items as opposed to a cost for non-produced items (i.e., items read silently). Recently, there have been two production effect studies that have explored the costs and benefits associated with reading aloud and reading silently; these studies have shown mixed results. For example, Bodner et al. (2014) used a within- versus between-subjects comparison to examine the costs and benefits associated with the production effect. They measured costs by comparing memory performance for silent items in within- versus between-subjects design and benefits by comparing memory performance for read aloud items in within- versus between-subjects design. Using discrimination (d’) as their dependent measure, Bodner et al. (2014) found no significant benefits for reading items aloud, but significant costs for reading items silently, suggesting that the production effect reflects the poor encoding or “lazy processing” of items read silently as opposed to the distinct encoding of items read aloud (see Hopkins & Edwards, 1972, for a similar explanation). Nevertheless, Bodner et al. (2014) also found greater memory performance for items read aloud in the within-subjects group compared to items read silently in the between-subjects group. They termed this “benefits-over-silent,” arguing that reading aloud does enhance memory for words regardless of study design.

Forrin et al. (2016) conducted a similar study to Bodner et al. (2014) but used hit rates as opposed to d’ scores in their analyses. Forrin et al. argued that d’ scores are a problematic measure for making comparisons between study design because false alarms have a different meaning in within versus between-subjects study designs. In a between-subject design, it is possible to calculate separate false alarm rates for items read aloud and items read silently because participants are only exposed to one production condition (either read aloud or read silently). Thus, a false alarm in the read aloud condition indicates that the participant mistakenly thought that he/she read the unstudied foil item aloud and a false alarm in the read silently condition indicates that the participant mistakenly thought that he/she read the unstudied foil item silently. In contrast, in a within-subjects design, foils are not normally distinguished by separate condition lists (e.g., a foil list for sing, a foil list for read aloud, a foil list for read silently), making it impossible to calculate separate false alarm rates. Thus, in Bodner et al. (2014), the d' scores used to calculate costs and benefits were calculated based on a common false alarm rate. This is especially problematic because when participants are asked to make study judgments at test (i.e., "Was this word studied aloud, silently, or not at all?"), they are more likely to exhibit a bias toward misclassifying a new foil word as read silently versus read aloud (e.g., Conway & Gathercole, 1987; Ozubko et al., 2012; Ozubko et al., 2014). The bias to misclassify foil words as read silently words versus read aloud words suggests that participants may have a more liberal response criterion when judging whether they previously studied a read silently item versus a read aloud item. Thus, it is important to calculate independent false alarm rates for each production condition (Sing, Read Aloud, Read Silently) so that they accurately represent *why* participants are responding 'yes' to an item that they did not study (e.g., is it because the participant thinks that he/she read it aloud at study or is it because the participant thinks that he/she read the item silently at study?).

Using the same analytic approach as Bodner et al. (2014) but with hit rates (as opposed to the more problematic d' scores), Forrin et al. (2016) found a significant benefit to reading aloud in a within- versus between-subjects design and a significant cost to reading silently in a within- versus between-subjects design; using d' scores, Forrin et al. (2016) found no significant costs or benefits. Forrin et al. (2016; p. 1102) argue that their findings of within- versus between-subjects benefits may have occurred because "the distinctiveness of aloud information was made salient at study, making it more accessible to participants at test than it was following a pure-aloud list."

Given that previous findings analyzing the costs and benefits associated with the production effect are mixed, and that singing is a novel form of production, it is valuable to explore the component processes that give rise to the greater production effect for singing versus reading aloud.

6.4.1 Costs and Benefits Analysis

To examine the costs and benefits associated with the production effect for singing, we analyzed the proportion hits (we used this dependent measure so that our findings could be easily compared to Forrin et al., 2016; see Appendix I for additional analyses using *A'* scores) for a within- and between-subjects experiment (Experiments 2 and 6, respectively[5]). The resulting analysis was a 3 x 2 mixed ANOVA with production condition (Sing, Read Aloud, Read Silently) as a within-subjects factor and study design (Within-Subjects, Between-Subjects) as a between-subjects factor. As shown in Figure 9, there was a significant main effect of production condition, $F(2, 76)=12.865$, $MSe=.013$, $p<.01$, ($\eta^2=.253$), indicating a robust overall production effect for Sing and Read Aloud conditions. There was not a significant main effect of

---

[5] The data from Experiment 2 was used for the current analysis because it involved the three production conditions of interest (i.e., Sing, Read Aloud, and Read Silently). In addition, Experiment 2 did not involve any additional manipulations, such as requiring participants with a known history of singing experience (Experiment 4) and asking participants to sing quickly (Experiment 5).

study design, *F*(1,38)=3.420, *MSe*=.040, *p*<.072, (η^2=.083). There was however, a significant interaction between condition and study design, *F*(2, 76)=12.865, *MSe*=.013, *p*<.01, (η^2=.242), supporting a significant production effect for Sing and Read Aloud conditions in the within-subjects design and no significant production effect for Sing and Read Aloud conditions in the between-subjects design (see Experiments 2 and 6 for analyses). The presence of a production effect for Sing and Read Aloud conditions in the within-subject design reflected a significant cost to reading silently in the within-subject design compared to the between-subjects design, *t*(38)= 3.985, *p*<.01. There was no significant benefit to singing or reading aloud in a within-subjects design compared to a between-subjects design (both ps>.450).

Taken together, these findings suggest that there is not a benefit for producing items via singing or reading aloud, but rather there seems to be a cost associated with not producing items (i.e., reading items silently). Moreover, conducting a similar analysis to Bodner et al. (2014), we examined "benefits-over-silent" by comparing hits for the Sing and Read Aloud conditions in the within-subjects design to hits for the Read Silently condition in the between-subjects design. In contrast to both Bodner et al. (2014) and Forrin et al. (2016), there were no significant "benefits-over-silent" for the Sing condition, *t*(19)= .893, *p*=.383, or the Read Aloud condition, *t*(19)= .999, *p*=.330.

In the following sections, our cost and benefits findings will be discussed in relation to the results of previous studies and theoretical accounts.

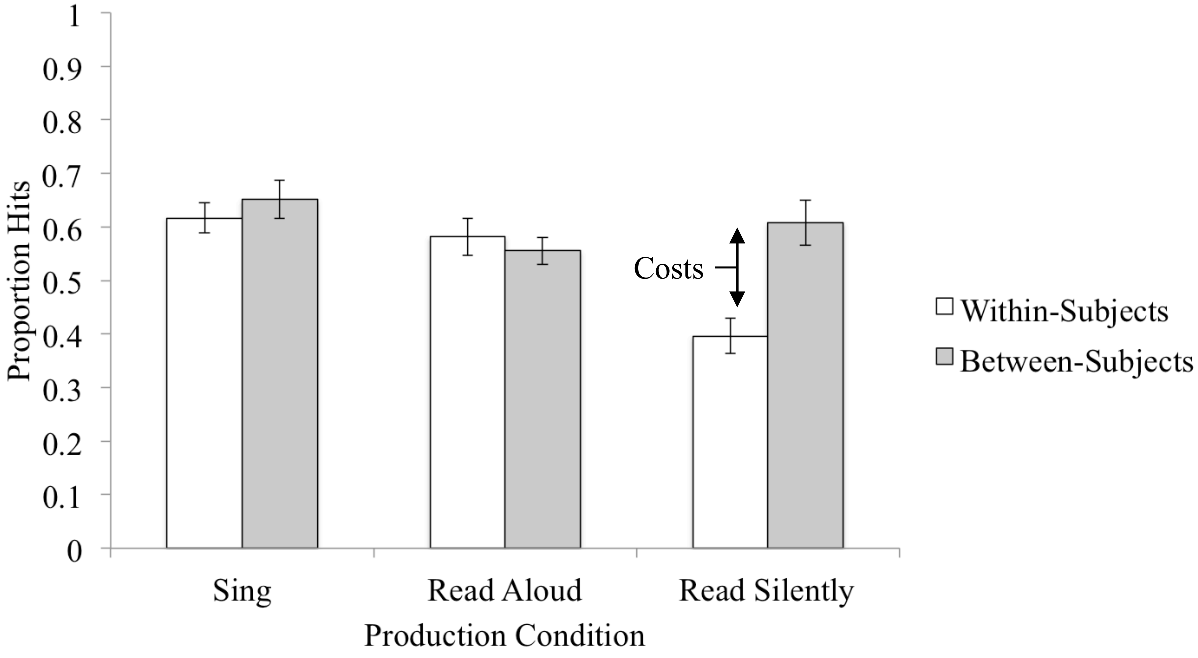


Figure 9. Proportion hits on the recognition test as a function of production condition (Sing, Read Aloud, Read Silently) and study design (Within, Between); error bars represent one standard error.

6.4.2 Relation to Theoretical Accounts of the Production Effect

Although it is generally assumed (and sometimes found; see Forrin et al., 2016) that benefits rather than costs underlie the production effect, this does not appear to be true in the current study. Our general findings are consistent with those of Bodner et al. (2014) in that there is a cost to reading items silently with no benefit for reading (or singing) items aloud. Given that we found no significant benefit for producing items (i.e., sing, read aloud), it is important to discuss these findings in relation to additional mechanisms beyond distinctiveness that may contribute to the production effect.

The notion that participants might not process the non-produced items as well as the produced items at study is directly related to the lazy reading hypothesis (Bodner et al., 2014; Forrin et al., 2016; see also Begg & Snider, 1987, as well as Begg & Roe, 1988). The lazy reading hypothesis suggests that memory performance is greater for produced than non-produced items in a within-subjects design because participants do not fully process the read silently items —rather they lazily read these items. The lazy reading of read silently items produces a significant production effect through decreased memory performance for unproduced items as opposed to increased memory performance for produced items.

For a number of reasons, we think it unlikely that the production effect can be explained solely by a lazy reading account. First, a recent study conducted by Hassall and colleagues (2016) closely followed the methodology of Experiment 2 of the current dissertation but incorporated electroencephalography (EEG) technology during the study phase, with the P300 measured relative to the onset of the production instruction. We found differences in the amplitude of the P300 component at encoding as a function of instruction (Sing, Read Aloud, Read Silently); there was a significantly greater P300 amplitude for produced versus non-produced items with no difference in the P300 amplitude for sing and read aloud items. Because the P300 component is thought to index cognitive processing (Comercheo & Polich, 1999; Donchin & Coles, 1988), these findings suggest that there are differences in the initial cognitive processing of items that are to be produced versus non-produced. At first glance, these findings seem to be consistent with a lazy reading hypothesis: greater cognitive processing in response to an instruction to produce rather than not produce. However, there was a P300 response to items in the Read Silently condition, which suggests that participants must have been engaging in some form of effortful cognitive processing (Kok, 2001) even in the absence of an overt production. If participants were simply lazily reading the read silently items, we would not have expected to observe a robust P300 response. Thus, the results of Hassall and colleagues (2016) suggest that the lazy reading of non-produced items is unlikely to entirely account for the findings of the production effect.

Second, the lazy reading of non-produced items cannot explain the occasionally reported presence of a blocked (Bodner et al., 2014) or between-subjects production effect (see Fawcett, 2013). The relative comparison of produced versus non-produced items is removed in blocked and between-subjects study designs, and thus any difference in memory performance cannot be driven by lazily processing the read silently items. In a standard between-subjects production study, participants are exposed to one of two conditions: Read Aloud or Read Silently. Because participants are aware that there is a memory test following the study phase, participants should be motivated to engage in effortful processing of the study items regardless of the condition (Read Aloud, Read Silently). The lazy reading hypothesis cannot account for the presence of a between-subjects production effect (Fawcett, 2013).

Third, if the production effect were simply due to the lazy reading of silently read items compared to produced items, there would not be any difference in the magnitude of the production effect as a function of the type of production. The lazy reading hypothesis would not predict that the lazy reading of items read silently would differentially affect memory performance for items in the other production conditions: Sing, Read Aloud Loudly, and Read Aloud. If lazy reading were the only mechanism underlying the production effect, we would not expect the magnitude of the effect to vary with type of production – lazy reading of the silent items should produce worse subsequent memory than that for *any* kind of active production. As evidenced by the findings of this dissertation, there were robust differences in the magnitude of the production effect for different types of production (i.e., Sing > Read Aloud Loudly > Read Aloud > Read Silently) that occurred consistently across five within-subjects experiments. The fact that singing produces a greater production effect, above and beyond that of reading aloud, strongly indicates that lazy reading cannot fully account for our pattern of results. Thus, while the lazy reading hypothesis may be able to partially account for the costs associated with the production effect, it cannot entirely account for the pattern of graded memory performance observed in the present dissertation.

6.4.3 Relation to the Distinctiveness Account

As just outlined above, the lazy reading hypothesis is unable to account for the findings of the present dissertation. Thus, it is important to examine the results of the cost/benefit analysis (Section 6.3.1) in relation to the distinctiveness account of the production effect. Despite our finding that the production effect appears to be more strongly associated with costs rather than benefits in within-subjects study designs, this finding is not necessarily inconsistent with a distinctiveness account. The distinctiveness account merely suggests that compared to non-produced items, produced items have at least one additional distinct element that is encoded at study and can be used to aid retrieval at test. Although one might assume that the encoding of this additional distinct element would lead to a benefit in memory performance for produced versus non-produced items, the fact that it does not result in a memory benefit does not necessarily contradict the distinctiveness account. In fact, the generation effect (Slamecka & Graf, 1978), which is also considered a form of distinct encoding, has also been associated with costs rather than benefits (Begg & Snider, 1987) and thought to reflect "… the inhibitory influence of the demand to generate on reading" (Begg & Snider, 1987, pp. 557). In addition, compared to between-subjects designs, within-subjects designs likely place a heavier cognitive load (i.e., greater cognitive demand) on participants (e.g., remembering what the coloured instruction means, carrying out the specific instruction) such that the items read silently may be relatively neglected because they do not require an active production response (see Jonides & Mack, 1984, for similar arguments in the context of accounting for the costs and benefits of attentional cueing).

Although it remains possible that other mechanisms beside distinctiveness may contribute (directly or indirectly) to the production effect, at present, there is no strong evidence that a mechanism other than distinctiveness contributes to the production either for reading aloud or for singing. We believe that the results of the current dissertation are best explained by a distinctiveness account. To date, there is no other theoretical explanation that can fully account for the graded association between the presumed quantity and quality of distinct elements encoded at study and subsequent memory performance at test. The finding of costs rather than benefits may be surprising but is not necessarily contradictory to a distinctiveness account.

6.5 APPLICATIONS

While our findings have clear implications for research investigating the production effect, they also have important implications for everyday life and, in particular, educational and clinical settings. Although the cost/benefit analysis suggests that the production effect for singing and read aloud may be attributed to costs rather than benefits, it is critical to remember that when the data from the within- and between-subjects study designs are combined, we continue to observe a significant production effect for the Sing and Read Aloud conditions. This suggests that regardless of the underlying component processes (costs versus benefits), the production is a true memory phenomenon that has the potential to be a valuable tool for educational and clinical settings.

6.5.1 Education

When studying for exams, students want to use effective memory strategies to remember information and boost their performance. The results of the current experiments suggest that various forms of vocal production, such as reading aloud, reading aloud loudly, and singing can have benefits on subsequent memory performance. Importantly, the present findings suggest that the most effective form of production is singing and it produces greater memory performance than reading aloud. Thus, singing could be a potentially effective memory strategy to use while studying for a quiz, test, and/or exam. One limitation to this possibility is that we used single study items (e.g., words such as “banana”, “cat”) in the current experiments, whereas longer sentences and full paragraphs are often studied within education settings. At present, we are unable to determine whether our findings using single words would extend to longer sentences and paragraphs of text, when singing is used as a form of production. Because Ozubko and colleagues (2012) found significantly greater memory performance for word pairs, sentences, and essays that were read aloud versus read silently, we might assume that the same may be true of singing. Importantly, Ozubko et al. (2012) found that the production effect can be extended from recognition memory tests to more educationally relevant tests, such as fill-in-the-blank statements (Experiment 3). This is a topic worthy of future investigation.

One limitation associated with using singing as a memory mnemonic for educational purposes is that it may not always be practical to sing aloud; thus the study environment of the individual would need to be one where they could (and would be allowed to) sing aloud. For example, it may not be appropriate to sing while studying in the library because there are other individuals trying to study quietly. Interestingly, Jamieson and Spear (2014) conducted a within-subjects production study wherein they asked participants to type words, imagine typing words, and read words silently. While they found that typing words produced greater memory performance than imagining typing words, they also found that both typing words and imagining typing words produced greater memory performance than reading words silently (i.e., Typing > Imagining Typing > Read Silently). The findings of Jamieson and Spear (2014) suggests that imagining singing words could be an effective memory mnemonic for students to use in educational studies —although imagining singing may not be as effective as singing aloud. In addition, singing may simply be awkward for many people even in the context of their own home. So, although singing has the potential to be an effective memory strategy in education settings, its utility likely depends upon the environmental context as well as the level of comfort associated with singing aloud.

6.5.2 Clinical

In addition to education settings, our findings have implications for work in clinical settings, especially for individuals diagnosed with a Neurocognitive Disorder. Neurocognitive Disorders involve impairment in "cognition that has not been present since birth or very early life, and thus represents a decline from a previously attained level of functioning" (American Psychiatric Association, 2013: Diagnostic and Statistical Manual of Mental Disorders -5th Edition [DSM-5], pp. 591). In the DSM-5, there are several etiological subtypes of Neurocognitive Disorders including (but not limited to) Alzheimer's disease, vascular, Lewy bodies, Parkinson's disease, Huntington's disease, frontotemporal, traumatic brain injury, and substance/medication-induced. One or more areas of cognitive functioning are affected in individuals who are diagnosed with Neurocognitive Disorders; these areas of cognitive functioning include attention, language, memory, spatial, and executive functions. In addition, there are different patterns of cognitive weaknesses and strengths depending on the etiology of the Neurocognitive Disorder (Wedding, 2007). For example, individuals who are diagnosed with a Neurocognitive Disorder due to Alzheimer's disease tend to exhibit learning and memory deficits with both verbal and visual information and across tests of recall and recognition. In contrast, individuals who are diagnosed with a Neurocognitive Disorder due to frontotemporal lobar degeneration tend to exhibit deficits in executive functioning, but learning and memory are relatively intact. For the purpose of the current dissertation, this section will focus entirely on Neurocognitive Disorders due to Alzheimer's disease (AD) because this etiological subtype involves the most pronounced deficits in learning and memory.

Although at present AD cannot be cured, the cognitive decline can be slowed. In the past, cholinesterase inhibitors, memantines , and neuroleptics have been several of the pharmacologic treatments used to slow the cognitive decline associated with AD. Although these medications have been shown to have short-term benefits (e.g., Ballard & Howard, 2006; Schneider, Dagerman, & Insel, 2006), the long-term benefits are less clear (e.g., Ballard, Margallo-Lana, Juszscak, Souglas, Swann, Thomas, et al., 2005; Schneider, Taiot, Dagerman, Davis, Hsiao, Ismail, et al., 2006). In particular, neuroleptics have also been associated with serious side effects including increased mortality (Schneider, Dagerman, & Insel, 2005), stroke (Schneider et al., 2006), and cerebrovascular and extrapyramidal symptoms (Ballard & Waite, 2006). Given the poor evidence for the effectiveness of pharmacological medication to slow the cognitive decline associated with AD, it is important to explore non-pharmacological interventions that can be effective in managing symptoms associated with AD. One form of non-pharmacological intervention that has shown benefits equal to or greater than pharmacological treatments is music therapy (e.g., Fossey, Ballard, Juszczak, James, Alder, Jacoby, et al., 2006).

Music therapy consists of the systematic use of musical instruments, dancing, and/or singing designed to increase positive feelings and motivation to improve the symptoms associated with AD (e.g., Goodall & Etters, 2005; Svansdottir & Snaedal, 2006). Music therapy has been shown to reduce the behavioural and social symptoms of AD (see Koger, Chapin, & Brotons, 1999 for review; also see Guetin, Portet, Picot, Pommie, Messaoudi, Djabelkir, Olsen, Cano, Lecourt, & Touchon, 2009). Following music therapy intervention, several studies have shown improvements in cognitive performance for individuals clinically diagnosed with AD (e.g., Särkämö, Tervaniemi, Laitinen, Numminen, Kurki, Johnson, & Rantanen, 2014). For instance, Särkämö and colleagues (2014) conducted a randomized controlled study, which randomly assigned 89 individuals diagnosed with AD to a 10-week singing coaching group, a 10-week listening coaching group, or a usual care control group. All participants completed neuropsychological assessment (including measures of cognition, mood, and quality of life) before the intervention, after the intervention, and 6 months following the intervention. Compared to the usual care control group, both the singing and music listening groups showed a significant improvement in overall cognitive performance. In addition, compared to the listening group, the singing group showed a significant increase in working memory abilities. These results suggest that music therapy can be an effective intervention in slowing and sometimes improving overall cognitive functioning, and that actively engaging in music via singing may have additional cognitive benefits above and beyond that of passively listening to music.

Despite the preliminary research of Särkämö and colleagues (2014), further research is needed to clarify the specific role of singing in cognition and memory, in particular. Although music and singing can have a positive impact on cognitive functioning in individuals who have AD (Särkämö et al., 2014), it would be interesting to examine whether there is a direct effect of singing a word or sentence on subsequent memory for that word or sentence in individuals diagnosed with AD. Such a study could use methodology similar to the current dissertation. Because music is a very unique and powerful therapeutic tool for individuals diagnosed with AD, it is quite possible that we would find a production effect for singing, especially given that singing involves so many of the distinct elements of music (e.g., intensity, pitch, and/or timbre; Roederer, 2008).

The clinical implications for such a study are valuable because easy, effective, and cost-efficient mnemonics are important for the quality of life of individuals diagnosed with Neurocognitive Disorders. For example, individuals who are in the early stages of AD often want to maintain quality of life by engaging in daily activities, such as self-care, grocery shopping, and attending social activities, but they often have difficulties remembering to shower daily; to do household chores; how to cook a meal; what to buy at the grocery store; and dates of social events. Thus, the utility of an effective mnemonic, that is easy to use and has no associated monetary cost, could improve quality of life for these individuals and their caregivers. For example, an individual diagnosed with a Mild Neurocognitive Disorder may find it helpful to sing or read aloud the steps of a recipe to aid memory for the sequence of steps involved in the recipe or the particular ingredient(s) to retrieve from the cupboard. In addition, individuals who are diagnosed with Mild Neurocognitive Disorders could use singing as a way to help facilitate the consolidation of tasks of daily living, such as medication instructions, phone numbers, and appointments.

6.6 CONCLUSION

This dissertation expands upon previous findings of the production effect while also aiding in our current understanding of the mechanisms underlying the production effect. The results of the current dissertation suggest that reading aloud in a normal voice is not necessarily the most advantageous form of production and that both reading aloud loudly and singing result in a greater production effect. In four within-subjects experiments, we consistently found evidence for a greater production effect for singing compared to reading aloud. Given our results in Experiments 4 through 6, we determined that our findings cannot be fully accounted for by bizarreness (e.g., Hirshman et al., 1989; Mather & Carstensen, 2005; Schmidt, 1991), a difference in production duration contributing to subsequent memory performance, or differences in elaborate encoding and trace memory strength (e.g., Bodner & Taikh, 2012). Rather, the most parsimonious explanation for our findings is a distinctiveness account of the production effect (e.g., Fawcett et al., 2012; Forrin et al., 2012; Forrin et al., 2016; Ozubko et al., 2013; Ozubko & MacLeod, 2010), which suggests that the number and type of potential distinct elements available at encoding is likely associated with subsequent memory performance at test and consequently with the magnitude of the production effect (i.e., the greater the number and type of distinct elements available at encoding, the greater the magnitude of the production effect). In addition to the number and type of distinct elements inherent in produced items, our findings suggest that the *relative within-subjects comparison* of these distinct elements is critical to the presence of the production; the removal of this relative comparison of items reduces the ability to use distinct elements as retrieval cues, thus eliminating the production effect.

BIBLIOGRAPHY

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th Edition). Arlington, VA: American Psychiatric Publishing.

Ballard, C., & Howard, R. (2006). Neuroleptic drugs in dementia: Benefits and harm. *Nature Reviews Neuroscience, 7*, 492-500.

Ballard, C., Margallo-Lana, M., Juszczak, E., Douglas, S., Swann, A., Thomas, A., et al. (2005). Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: Randomized double blind placebo controlled trial. *British Medical Journal, 330*, 874-878.

Ballard, C., & Waite, J. (2006). The effectiveness of atypical antipsychotics for the treatment of aggression and psychosis in Alzheimer's disease. *Cochrane Database Systematic Reviews, 25*, CD003476.

Baur, B., Uttner, I., Ilmberger, J., Fesl, G., & Mai, N. (2000). Music memory provides access to verbal knowledge in a patient with global amnesia. *Neurocase, 6*, 415-421.

Begg, I., & Roe, H. (1988). On the inhibition of reading by generation. *Canadian Journal of Psychology, 42*, 325-336.

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 553-563.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*, 201-210.

Bettison, S. (1996). The long-term effects of auditory training on children with autism. *Journal of Autism and Developmental Disorders, 26*, 361-374.

Bodner, G. E., & MacLeod, C. M. (2016). The benefits of studying by production… and of studying production: Introduction to the special issue on the production effect in memory. *Canadian Journal of Experimental Psychology, 70*, 89-92.

Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1711-1719.

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review, 21*, 149-154.

Castel, A. D. (November, 2009). *Predicting the production effect: The use and misuse of self-generated cues*. Paper presented at the 5th bi-annual meeting of the International Association on Metacognition, Boston, MA.

Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition*, *41*, 28-35.

Claussen, D. W., & Thaut, M. H. (1997). Music as a mnemonic device for children with learning disabilities. *Canadian Journal of Music Therapy, 5*, 55-66.

Cohen, R. L. (1981) "On the generality of some memory laws". *Scandinavian Journal of Psychology*, *22,* 267-281.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, and Computers, 25*, 257–271.

Camerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology, 110*, 24-30.

Conway, M. A., & Gatherole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language, 26*, 341-361.

Cooper, E. C., & Pantle, A. J. (1967). The total-time hypothesis in verbal learning. *Psychological Bulletin, 68*, 221-234.

Craik, F. I. M. (2002). Levels of processing: Past, present…and future? *Memory, 10*, 305-318.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671-684.

Dodson, C. S. & Schacter, D. L. (2001). “If I had said it I would have remembered it”: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletic & Review*, *8*, 155-161.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523-533.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11, 357-374.

Eich, E., & Metcalfe, J. (1989). Mood-dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 443-455.

El Haj, M., Fasotti, L., & Allain, P. (2012). The involuntary nature of music-evoked autobiographical memories in Alzheimer's disease. *Consciousness and Cognition, 21,* 238-246.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 671-682.

Engelkamp, J., & Krumnacker, H. (1980). Image- and motor-processes in the retention of verbal materials. *für Experimentelle und Angewandte Psychologie, 27*, 511-533.

Epstein, C. M., Meador, K. J., Loring, D. W., Wright, R. J., Weissman, J. D., Sheppard, S., Lah, J. J., Puhalovich, F., Gaitan, L., & Davey, K. R. (1999). Localization and characterization of speech arrest during transcranial magnetic stimulation. *Clinical Neurophysiology, 110*, 1073-1079.

Fawcett, J. M. (2013). The production effect benefits performance in between-subjects designs: A meta-analysis. *Acta Psychologica, 142*, 1-5.

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology, 70*, 99-115.

Fawcett, J. M., Quinlan, C. K., & Taylor T. L. (2012). Interplay of the production and picture superiority effect: A signal detection analysis. *Memory*, *20*, 655-666.

Forrin, N. D., Groot, B., & MacLeod, C. M. (2016). The d-Prime directive: Assessing costs and benefits in recognition by dissociating mixed-list false alarm rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1090-1111.

Forrin, N. D., & MacLeod, C. M. (2016). Auditory presentation at test does not diminish the production effect in recognition. *Canadian Journal of Experimental Psychology, 70*, 116-124.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition, 40*, 1046-1055.

Fossey, J., Ballard, C., Juszczak, E., James, I., Alder, N., Jacoby, R., et al. (2006). Effect of enhanced psychosocial care on antipsychotic use in nursing home residents with severe dementia: Cluster randomised trial. *British Medical Journal, 332*, 756-761.

Foster, N. A., & Valentine, E. R. (2001). The effect of auditory stimulation on autobiographical recall in dementia. *Experimental Aging Research, 27*, 215-228.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6*.

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition, 16*, 110-119.

Gfeller, K. E. (1983). Musical mnemonics as an aid to retention with normal and learning-disabled students. *Journal of Music Therapy, 20*, 179-189.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*, 325-331.

Goodall, D., & Etters, L. (2005). The therapeutic use of music on agitated behavior in those with dementia. *Holistic Nursing Practice, 19*, 258-262.

Goodwin, D. W., Powell, B., Bremer, D., Hoine, H., & Stern, J. (1969). Alcohol and recall: State dependent effects in man. *Science, 163*, 1358.

Guetin, S., Portet, F., Picot, M. C., Pommie, C., Messaoudi, M., Djabelkir, L., Olsen, A. L., Cano, M. M., Lecourt, E., & Touchon, J. (2009). Effect of music therapy on anxiety and depression in patients with Alzheimer's type dementia: Randomized, controlled study. *Dementia and Geriatric Cognitive Disorders, 28*, 36-46.

Haslam, C., & Cook, M. (2002). Striking a chord with amnesic patients: Evidence that song facilitates memory. *Neurocase, 8,* 453-465.

Hassall, C., Quinlan, C. K., Turk, D.J., Taylor, T. L., & Krigolson, O. (2016). A preliminary investigation into the neural basis of the production effect. *Canadian Journal of Experimental Psychology, 70*, 139-146.

Hirshman, E., Whelley, M. M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory and Language, 28,* 594-609.

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior, 11*, 534-537.

Hopkins, R. H., Boylan, R. J., & Lincoln, G. L. (1972). Pronunciation and apparent frequency. *Journal of Verbal Learning and Verbal Behavior, 11*, 105-113.

Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology, 62,* 242-246.

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2*, 105-112.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). New York, NY: Oxford University Press.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421-445.

Irish, M., Cunningham, C. J., Walsh, J. B., Coakly, D., Lawlor, A. B., Robertson, I. H., et al (2006). Investigating the enhancing effect of music on autobiographical memory in mild Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders, 22*, 108-120.

Jamieson, R. K., & Spear, J. (2014). The offline production effect. *Canadian Journal of Experimental Psychology*, 68, 20-28.

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology, 70*, 154-164.

Jeffries, K.J., Fritz, J. B., & Braun, A. R. (2003). Word in melody: an $H_2{}^{15}O$ PET study of brain activation during singing and speaking. *NeuroReport, 14*, 749-754.

Jones, A. C., & Pyc, M. A (2013). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2,* 1-6.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96*, 29-44.

Jonker, T. R., Levene, M., & MacLeod, C. M. (2013). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 441-448.

Koger, S. M., Chapin, K., & Brotons, M. (1999). Is music therapy an effective intervention for dementia? A meta-analytic review of literature. *Journal of Music Therapy, 36*, 2-15.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38, 557–577.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. (1976). Depth of processing, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 75-102). Oxford, England: Wiley.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98,* 291-310.

MacLeod, C. M. (1996). How priming affects two speeded implicit tests of remembering: Naming colors versus reading words. *Consciousness & Cognition, 5*, 73-90.

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review, 18,* 1197-1202.

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26,* 390-395.

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*, 671-685.

MacLeod, C. M., & Masson, M. E. J. (1997). Enhancement patterns are different in masked word identification and word fragment completion. *Journal of Memory and Language, 36*, 461-483.

Macmillan, N., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum.

Mather, M., & Carstensen, L. L., (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences, 9*, 496-502.

Matlin, M. W. (2005). *Cognition*. Crawfordsville: John Wiley & Sons, Inc.

McDaniel, M. A., Anderson, D. C., Einstein, G. O., & O'Halloran, C. M. (1989). Modulation of environmental reinstatement effects through encoding strategies. *American Journal of Psychology, 102*, 523-548.

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15*, 237-255.

McDaniel, M. A., & Einstein, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12*, 54-65.

McDaniel, M. A., Einstein, G. O., De Losh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 422-435.

Murdock, B. B., Jr. (1960). The distinctiveness of stimuli. *Psychological Review, 67*, 16–31.

Oostendorp, J. C., & Montel, S. R. (2014). Singing can enhance episodic memory functioning in elderly people with Alzheimer's disease. *Journal of the American Geriatrics Society, 62*, 982-983.

Ozdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage, 33*, 628-635.

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition, 40*, 326-338.

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and benefits memory for text. *Memory*, *20*, 717-727.

Ozubko, J. D. & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36,* 1543-1547.

Ozubko, J.D., Major, J., & MacLeod, C.M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory, 22*, 509-524.

Paivio, A. (1971). *Imagery and verbal processes.* Oxford, UK: Holt, Rinehart & Winston.

Palmer, I. (unpublished). *Duration of production as a factor in enhanced recognition on a memory task: A study of the production effect*.

Peters, R., & McGee, R. (1982). Cigarette smoking and state-dependent memory. *Psychopharmacology, 76*, 232-235.

Quinlan, C. K., & Fawcett, J. M. (unpublished). *The production effect: A meta-analytic review*.

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*, 904-915.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013). How does using object names influence visual recognition memory? *Journal of Memory and Language, 68*, 10-25.

Roederer, J. G. (2008). *The physics and psychophysics of music: An introduction* (pp. 1-242). New York, NY: Springer.

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

Särkämö, T., Tervaniemi, M., Laitinen, S., Numminen, A., Kurki, M., Johnson, J. K., & Rantanen, P. (2014). Cognitive, emotional, and social benefits of regular musical activities in early dementia: Randomized controlled study. *The Gerontologist, 54*, 634-650.

Schacter, D. L. Israel, L. & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1-24.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19,* 523-542.

Schneider, L. S., Dagerman, K. S., & Insel, P. (2005). Risk of death with atypical antipsychotic drug treatment for dementia: Meta-analysis of randomized placebo-controlled trials. *Journal of the American Medical Association, 294*, 1934-1943.

Schneider, L. S., Dagerman, K., & Insel, P. S. (2006). Efficacy and adverse effects of atypical antipsychotics for dementia: Meta-analysis of randomized, placebo-controlled trials. *American Journal of Geriatric Psychology, 14*, 191-210.

Schneider, L. S., Tariot, P. N., Dagerman, K. S., Davis, S. M., Hsiao, J. K., Ismail, S. M., et al. (2006). Effectiveness of atypical antipsychotic drugs in patients with Alzheimer's disease. *The New England Journal of Medicine, 355*, 1525-1538.

Simmons-Stern, N. R., Budson, A. E., & Ally, B. A. (2010). Music as a memory enhancer in patients with Alzheimer's disease. *Neuropsychologia, 48*, 3164-3167.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592-604.

Sternberg, R. J. (1999). *Cognitive psychology (2nd Edition)*. Fort Worth, TX: Harcourt Brace College Publishers.

Stewart, L.,Walsh, V., Frith, U., & Rothwell, J. C. (2001). Transcranial magnetic stimulation produces speech arrest but not song arrest. *Annals of the New York Academy of Sciences, 930*, 433-435.

Svansdottir, H. B., & Snaedal, J. (2006). Music therapy in moderate and severe dementia of Alzheimer's type: A case-control study. *International Psychogeriatrics, 18*, 613-621.

Taikh, A., & Bodner, G. E. (2016). Evaluating the basis of the between-group production effect in recognition. *Canadian Journal of Experimental Psychology, 70*, 186-194.

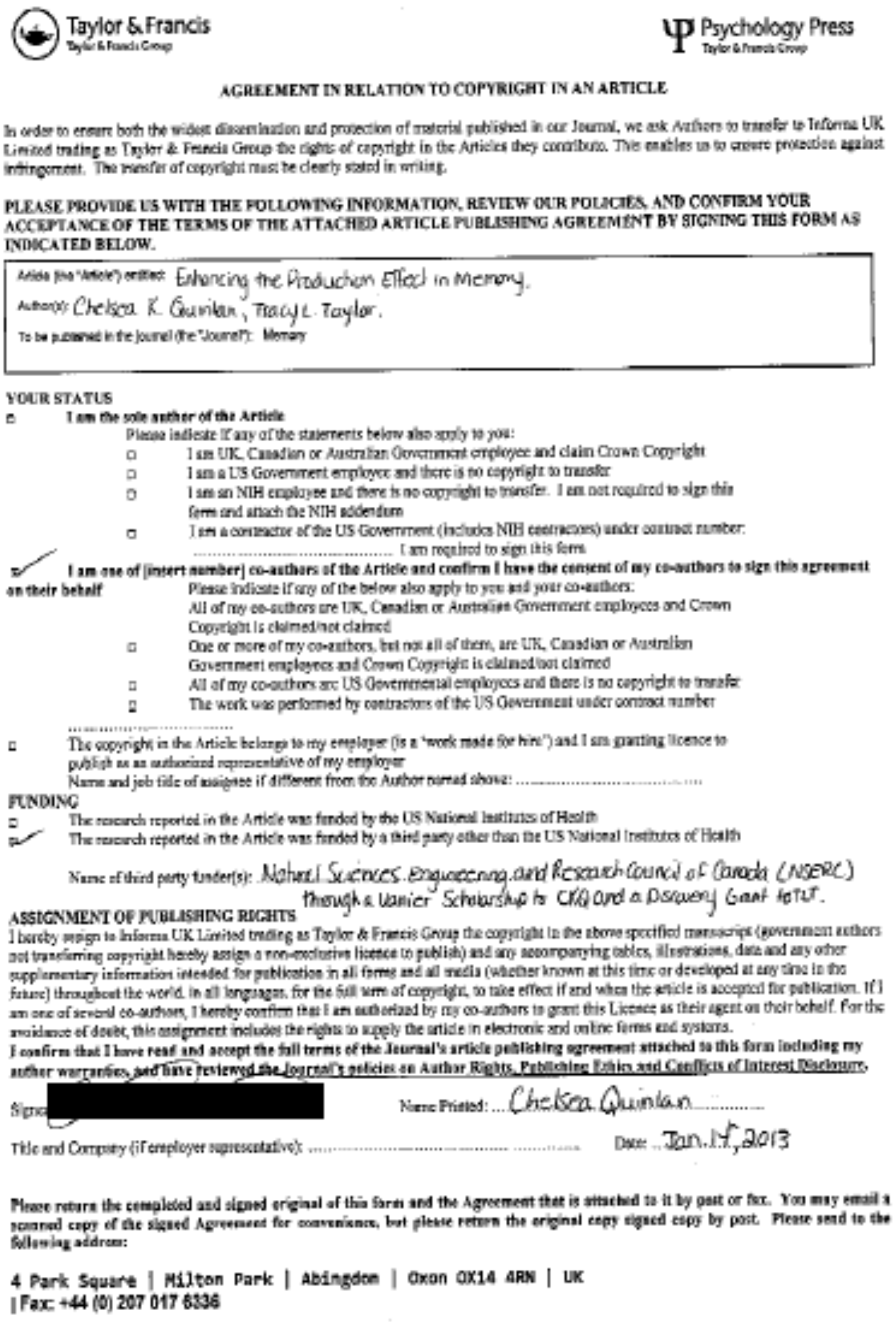
Thaut, M. H., Peterson, D. A., & McIntosh, G. C. (2005). Musical mnemonics induce brain plasticity and oscillatory synchrony in neural networks underlying memory. *Annals of the New York Academy of Sciences, 1060,* 243-254.

Wedding, D. (2008). The neuropsychology handbook (3rd edition). New York, NY: Springer Publishing Company.

Wilde, O. (1895). The importance of being earnest. pp. 1-67.

Wolfe, D. E., & Hom, C. (1993). Use of melodies as structural prompts for learning and retention of sequential verbal information by preschool students. *Journal of Music Therapy, 30*, 100-118.

APPENDIX A COPYRIGHT AGREEMENT AND PERMISSIONS



Taylor & Francis
Taylor & Francis Group

Psychology Press
Taylor & Francis Group

AGREEMENT IN RELATION TO COPYRIGHT IN AN ARTICLE

In order to ensure both the widest dissemination and protection of material published in our Journal, we ask Authors to transfer to Informa UK Limited trading as Taylor & Francis Group the rights of copyright in the Articles they contribute. This enables us to ensure protection against infringement. The transfer of copyright must be clearly stated in writing.

PLEASE PROVIDE US WITH THE FOLLOWING INFORMATION, REVIEW OUR POLICIES, AND CONFIRM YOUR ACCEPTANCE OF THE TERMS OF THE ATTACHED ARTICLE PUBLISHING AGREEMENT BY SIGNING THIS FORM AS INDICATED BELOW.

Article (the "Article") entitled: Enhancing the Production Effect in Memory.
Author(s): Chelsea K. Quinlan, Tracy L. Taylor.
To be published in the journal (the "Journal"): Memory

YOUR STATUS

☐ **I am the sole author of the Article**
Please indicate if any of the statements below also apply to you:
- ☐ I am UK, Canadian or Australian Government employee and claim Crown Copyright
- ☐ I am a US Government employee and there is no copyright to transfer
- ☐ I am an NIH employee and there is no copyright to transfer. I am not required to sign this form and attach the NIH addendum
- ☐ I am a contractor of the US Government (includes NIH contractors) under contract number: .... I am required to sign this form

☑ **I am one of [insert number] co-authors of the Article and confirm I have the consent of my co-authors to sign this agreement on their behalf**
Please indicate if any of the below also apply to you and your co-authors:
All of my co-authors are UK, Canadian or Australian Government employees and Crown Copyright is claimed/not claimed
- ☐ One or more of my co-authors, but not all of them, are UK, Canadian or Australian Government employees and Crown Copyright is claimed/not claimed
- ☐ All of my co-authors are US Governmental employees and there is no copyright to transfer
- ☐ The work was performed by contractors of the US Government under contract number ....

☐ The copyright in the Article belongs to my employer (is a "work made for hire") and I am granting licence to publish as an authorized representative of my employer
Name and job title of assignee if different from the Author named above: ....

FUNDING

☐ The research reported in the Article was funded by the US National Institutes of Health
☑ The research reported in the Article was funded by a third party other than the US National Institutes of Health

Name of third party funder(s): Natural Sciences Engineering and Research Council of Canada (NSERC) through a Vanier Scholarship to CKQ and a Discovery Grant to TLT.

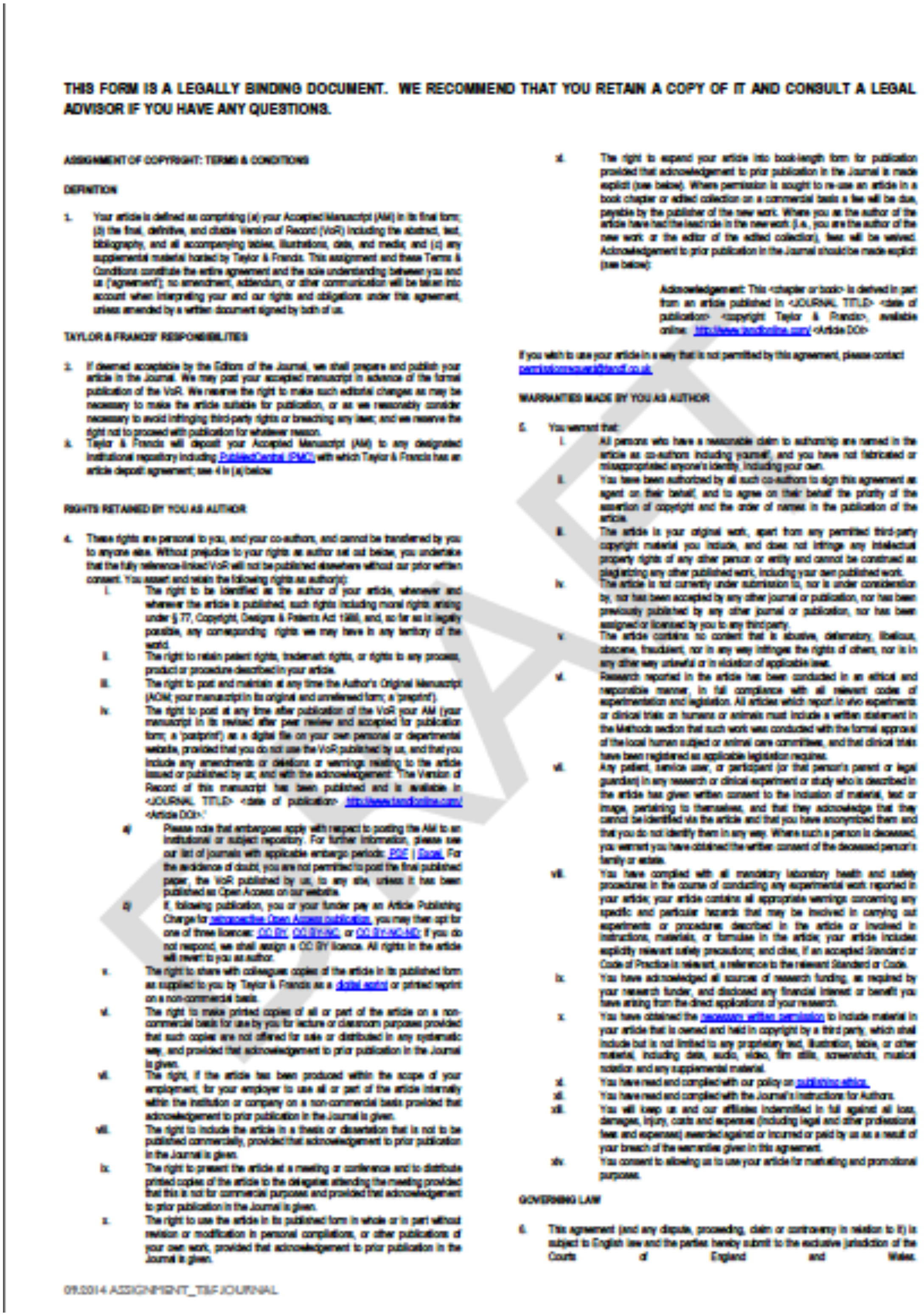
ASSIGNMENT OF PUBLISHING RIGHTS

I hereby assign to Informa UK Limited trading as Taylor & Francis Group the copyright in the above specified manuscript (government authors not transferring copyright hereby assign a non-exclusive licence to publish) and any accompanying tables, illustrations, data and any other supplementary information intended for publication in all forms and all media (whether known at this time or developed at any time in the future) throughout the world, in all languages, for the full term of copyright, to take effect if and when the article is accepted for publication. If I am one of several co-authors, I hereby confirm that I am authorized by my co-authors to grant this Licence as their agent on their behalf. For the avoidance of doubt, this assignment includes the rights to supply the article in electronic and online forms and systems.
I confirm that I have read and accept the full terms of the Journal's article publishing agreement attached to this form including my author warranties, and have reviewed the Journal's policies on Author Rights, Publishing Ethics and Conflicts of Interest Disclosure.

Sign [redacted] Name Printed: Chelsea Quinlan
Title and Company (if employer representative): .... Date: Jan. 14, 2013

Please return the completed and signed original of this form and the Agreement that is attached to it by post or fax. You may email a scanned copy of the signed Agreement for convenience, but please return the original copy signed copy by post. Please send to the following address:

4 Park Square | Milton Park | Abingdon | Oxon OX14 4RN | UK
| Fax: +44 (0) 207 017 6336



THIS FORM IS A LEGALLY BINDING DOCUMENT. WE RECOMMEND THAT YOU RETAIN A COPY OF IT AND CONSULT A LEGAL ADVISOR IF YOU HAVE ANY QUESTIONS.

ASSIGNMENT OF COPYRIGHT: TERMS & CONDITIONS

DEFINITION

1. Your article is defined as comprising (a) your Accepted Manuscript (AM) in its final form; (b) the final, definitive, and citable Version of Record (VoR) including the abstract, text, bibliography, and all accompanying tables, illustrations, data, and media; and (c) any supplemental material hosted by Taylor & Francis. This assignment and these Terms & Conditions constitute the entire agreement and the sole understanding between you and us ('agreement'); no amendment, addendum, or other communication will be taken into account when interpreting your and our rights and obligations under this agreement, unless amended by a written document signed by both of us.

TAYLOR & FRANCIS' RESPONSIBILITIES

2. If deemed acceptable by the Editors of the Journal, we shall prepare and publish your article in the Journal. We may post your accepted manuscript in advance of the formal publication of the VoR. We reserve the right to make such editorial changes as may be necessary to make the article suitable for publication, or as we reasonably consider necessary to avoid infringing third-party rights or breaching any laws; and we reserve the right not to proceed with publication for whatever reason.
3. Taylor & Francis will deposit your Accepted Manuscript (AM) to any designated institutional repository including PubMedCentral (PMC) with which Taylor & Francis has an article deposit agreement; see 4 iv (a) below.

RIGHTS RETAINED BY YOU AS AUTHOR

4. These rights are personal to you, and your co-authors, and cannot be transferred by you to anyone else. Without prejudice to your rights as author set out below, you undertake that the fully reference-linked VoR will not be published elsewhere without our prior written consent. You assert and retain the following rights as author(s):
   - i. The right to be identified as the author of your article, whenever and wherever the article is published, such rights including moral rights arising under § 77, Copyright, Designs & Patents Act 1988, and, so far as is legally possible, any corresponding rights we may have in any territory of the world.
   - ii. The right to retain patent rights, trademark rights, or rights to any process, product or procedure described in your article.
   - iii. The right to post and maintain at any time the Author's Original Manuscript (AOM; your manuscript in its original and unrefereed form; a 'preprint').
   - iv. The right to post at any time after publication of the VoR your AM (your manuscript in its revised after peer review and accepted for publication form; a 'postprint') as a digital file on your own personal or departmental website, provided that you do not use the VoR published by us, and that you include any amendments or deletions or warnings relating to the article issued or published by us; and with the acknowledgement: 'The Version of Record of this manuscript has been published and is available in <JOURNAL TITLE> <date of publication> http://www.tandfonline.com/ <Article DOI>.'
     - a) Please note that embargoes apply with respect to posting the AM to an institutional or subject repository. For further information, please see our list of journals with applicable embargo periods: PDF | Excel. For the avoidance of doubt, you are not permitted to post the final published paper, the VoR published by us, to any site, unless it has been published as Open Access on our website.
     - b) If, following publication, you or your funder pay an Article Publishing Charge for retrospective Open Access publication, you may then opt for one of three licences: CC BY, CC BY-NC, or CC BY-NC-ND; if you do not respond, we shall assign a CC BY licence. All rights in the article will revert to you as author.
   - v. The right to share with colleagues copies of the article in its published form as supplied to you by Taylor & Francis as a digital eprint or printed reprint on a non-commercial basis.
   - vi. The right to make printed copies of all or part of the article on a non-commercial basis for use by you for lecture or classroom purposes provided that such copies are not offered for sale or distributed in any systematic way, and provided that acknowledgement to prior publication in the Journal is given.
   - vii. The right, if the article has been produced within the scope of your employment, for your employer to use all or part of the article internally within the institution or company on a non-commercial basis provided that acknowledgement to prior publication in the Journal is given.
   - viii. The right to include the article in a thesis or dissertation that is not to be published commercially, provided that acknowledgement to prior publication in the Journal is given.
   - ix. The right to present the article at a meeting or conference and to distribute printed copies of the article to the delegates attending the meeting provided that this is not for commercial purposes and provided that acknowledgement to prior publication in the Journal is given.
   - x. The right to use the article in its published form in whole or in part without revision or modification in personal compilations, or other publications of your own work, provided that acknowledgement to prior publication in the Journal is given.
   - xi. The right to expand your article into book-length form for publication provided that acknowledgement to prior publication in the Journal is made explicit (see below). Where permission is sought to re-use an article in a book chapter or edited collection on a commercial basis a fee will be due, payable by the publisher of the new work. Where you as the author of the article have had the lead role in the new work (i.e., you are the author of the new work or the editor of the edited collection), fees will be waived. Acknowledgement to prior publication in the Journal should be made explicit (see below):

     > Acknowledgement: This <chapter or book> is derived in part from an article published in <JOURNAL TITLE> <date of publication> <copyright Taylor & Francis>, available online: http://www.tandfonline.com/ <Article DOI>

If you wish to use your article in a way that is not permitted by this agreement, please contact permissionrequest@tandf.co.uk

WARRANTIES MADE BY YOU AS AUTHOR

5. You warrant that:
   - i. All persons who have a reasonable claim to authorship are named in the article as co-authors including yourself, and you have not fabricated or misappropriated anyone's identity, including your own.
   - ii. You have been authorized by all such co-authors to sign this agreement as agent on their behalf, and to agree on their behalf the priority of the assertion of copyright and the order of names in the publication of the article.
   - iii. The article is your original work, apart from any permitted third-party copyright material you include, and does not infringe any intellectual property rights of any other person or entity and cannot be construed as plagiarizing any other published work, including your own published work.
   - iv. The article is not currently under submission to, nor is under consideration by, nor has been accepted by any other journal or publication, nor has been previously published by any other journal or publication, nor has been assigned or licensed by you to any third party.
   - v. The article contains no content that is abusive, defamatory, libellous, obscene, fraudulent, nor in any way infringes the rights of others, nor is in any other way unlawful or in violation of applicable laws.
   - vi. Research reported in the article has been conducted in an ethical and responsible manner, in full compliance with all relevant codes of experimentation and legislation. All articles which report in vivo experiments or clinical trials on humans or animals must include a written statement in the Methods section that such work was conducted with the formal approval of the local human subject or animal care committees, and that clinical trials have been registered as applicable legislation requires.
   - vii. Any patient, service user, or participant (or that person's parent or legal guardian) in any research or clinical experiment or study who is described in the article has given written consent to the inclusion of material, text or image, pertaining to themselves, and that they acknowledge that they cannot be identified via the article and that you have anonymized them and that you do not identify them in any way. Where such a person is deceased, you warrant you have obtained the written consent of the deceased person's family or estate.
   - viii. You have complied with all mandatory laboratory health and safety procedures in the course of conducting any experimental work reported in your article; your article contains all appropriate warnings concerning any specific and particular hazards that may be involved in carrying out experiments or procedures described in the article or involved in instructions, materials, or formulae in the article; your article includes explicitly relevant safety precautions; and cites, if an accepted Standard or Code of Practice is relevant, a reference to the relevant Standard or Code.
   - ix. You have acknowledged all sources of research funding, as required by your research funder, and disclosed any financial interest or benefit you have arising from the direct applications of your research.
   - x. You have obtained the necessary written permission to include material in your article that is owned and held in copyright by a third party, which shall include but is not limited to any proprietary text, illustration, table, or other material, including data, audio, video, film stills, screenshots, musical notation and any supplemental material.
   - xi. You have read and complied with our policy on publishing ethics.
   - xii. You have read and complied with the Journal's Instructions for Authors.
   - xiii. You will keep us and our affiliates indemnified in full against all loss, damages, injury, costs and expenses (including legal and other professional fees and expenses) awarded against or incurred or paid by us as a result of your breach of the warranties given in this agreement.
   - xiv. You consent to allowing us to use your article for marketing and promotional purposes.

GOVERNING LAW

6. This agreement (and any dispute, proceeding, claim or controversy in relation to it) is subject to English law and the parties hereby submit to the exclusive jurisdiction of the Courts of England and Wales.

09.2014 ASSIGNMENT_T&F JOURNAL

APPENDIX B EXPERIMENT 1: EXAMPLE OF EXPERIMENT INSTRUCTIONS

Overview:

This experiment will present you with a series of study trials followed by a memory task.

Study Trials:

At the start of each trial, a blank screen will be present for a short duration.
Then, a word will appear in one of three colours.
If the word is presented in BLUE, you should read that word aloud loudly.
If the word is presented in RED, you should read that word aloud.
If the word is presented in WHITE, you should read that word silently (i.e., not aloud).

Memory Task:

As you go through the study trials, you should be trying to remember the words.

At the end of the study trials, you will be given a memory task. The instructions for this task will be up on the screen when it is time to do it.

Before you Begin:

Before you begin, you will receive 15 exposures to the BLUE (Sing), RED (Read Aloud), and WHITE (Read Silently) instructions so that you may become familiar with them.

After these 15 exposures, you will receive 15 practice trials where you will follow the colour instructions. In this phase, if you see the word in BLUE, read the word aloud loudly; if you see the word in RED, read the word aloud; and if you see the word in WHITE, read the word silently.

The experiment will begin following the end of these 15 practice trials.

If you have any questions, please address them now to the experimenter.

APPENDIX C ADDITIONAL ANALYSES FOR EXPERIMENT 1

Mean proportions

The mean proportion of yes responses were analyzed in a one-way ANOVA with condition (Read Aloud Loudly, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, $F(2, 26)=45.502$, $MSe=.006$, $p<.001$, ($\eta^2=.778$). All contrasts were significant, p*'s* $<.001$.

An alternative measure of sensitivity: d'

The d' scores, which are another measure of sensitivity that takes into account hits and false alarm rates, were analyzed in a one-way ANOVA with condition (Read Aloud Loudly, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, $F(2, 26)=20.021$, $MSe=.038$, $p<.001$, ($\eta^2=.606$). As with the mean proportion of yes responses, all contrasts were significant, p's $<.001$.

Response bias

A' is often accompanied by B''_D, which is a nonparametric measure of response bias. Lower B''_D values represent more liberal responding and higher B''_D values represent more conservative responding. For the interested reader (also see Fawcett, Quinlan, & Taylor, 2012), the B''_D scores were analyzed in a one-way ANOVA with condition (Read Aloud Loudly, Read Aloud, Read Silently) as a within-subjects factor. There was a significant effect of condition, $F(2, 26)=10.195$, $MSe=.083$, $p=.001$, ($\eta^2=.440$). Contrasts showed that the B''_D scores were significantly lower in the Read Aloud Loudly condition ($M=.028$, $SD=.615$) than in the Read Aloud condition ($M=.225$, $SD=.507$; $t(13)=2.864$, $p=.013$), as well as the Read Silently condition, ($M=.517$, $SD=.388$; $t(13)=3.539$, $p=.004$. Also, B''_D scores were significantly lower in the Read Aloud condition than in the Read Silently condition, $t(13)=2.689$, $p=.019$.

APPENDIX D ADDITIONAL ANALYSES FOR EXPERIMENT 2

Mean proportions

The mean proportions of yes responses were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, $F(2, 38)=43.333$, $MSe=.008$, $p<.001$, ($\eta^2=.695$). All contrasts were significant, *p's* <.001.

An alternative measure of sensitivity: d'

The *d'* scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, $F(2, 38)=16.813$, $MSe=.094$, $p<.001$, ($\eta^2=.469$). As with the mean proportion of yes responses, all contrasts were significant, *p*'s <.001.

Response bias

The B''_D scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant effect of condition, $F(2, 38)=12.680$, $MSe=.047$, $p<.001$, ($\eta^2=.400$). Contrasts showed that the B''_D scores were significantly lower in both the Sing condition ($M=.329$, $SD=.476$) and the Read Aloud condition ($M=.405$, $SD=.462$) than in the Read Silently condition ($M=.659$, $SD=.446$; $t(19)=4.894$, $p<.001$, and $t(19)=3.693$, $p=.002$, respectively). There was no significant difference in the B''_D scores for the Sing condition compared to the Read Aloud condition, $t(19)=1.091$, $p=.289$.

APPENDIX E ADDITIONAL ANALYSES FOR EXPERIMENT 3

Mean proportions

The mean proportions of yes responses were analyzed in a one-way ANOVA with condition (Sing, Read Aloud Loudly, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, *F*(3, 63)=22.529, *MSe*=.009, *p*<.001, (η^2=.518). All of the contrasts were significant, all *p's* <.01, with the exception of the comparison between the Read Aloud Loudly (*M*=.533; *SD*=.156) and Read Aloud conditions (*M*=.543; *SD*=.139; *t*(21)=.417, *p*=.681).

An alternative measure of sensitivity: d'

The *d'* scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud Loudly, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, *F*(3, 63)=12.725, *MSe*=.120, *p*<.001, (η^2=.377). Contrasts showed that the *d'* scores were significantly greater in the Sing condition (*M*=1.180, *SD*=.381) compared to the Read Aloud Loudly condition (*M*=1.001, *SD*=.494; *t*(21)=2.262, *p*=.034), the Read Aloud condition (*M*=.673, *SD*=.352; *t*(21)=5.604, *p*<.001), and the Read Silently condition, (*M*=.631, *SD*=.404; *t*(21)=4.948, *p*<.001). The *d'* scores were also significantly greater in the Read Aloud Loudly condition compared to the Read Aloud condition, *t*(21)=3.104, *p*=.005, and the Read Silently condition, *t*(21)=3.010, *p*=.007. However, the *d'* scores were not significantly different for the Read Aloud and Read Silently conditions, *t*(21)=.373, *p*=.713).

Response bias

The B''_D scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud Loudly, Read Aloud, Read Silently) as a within-subjects factor. There was a significant effect of condition, *F*(3, 63)=9.972, *MSe*=.066, *p*<.001, (η^2=.322). Contrasts showed that the B''_D scores were significantly greater in the Read Silently condition (*M*=.658, *SD*=.328) than in the Read Aloud condition (*M*=.283, *SD*=.453; *t*(21)=5.752, *p*<.001); the Read Aloud Loudly condition (*M*=.447, *SD*=.482; *t*(21)=3.250, *p*=.004); and the Sing condition (*M*=.301, *SD*=.555; *t*(21)=5.063, *p*<.001). The B''_D scores were also significantly lower in the Read Aloud condition compared to the Read Aloud Loudly condition, *t*(21)=2.580, *p*=.017). There were no significant differences in B''_D scores for the Sing condition compared to the Read Aloud Loudly condition, *t*(21)=1.509, *p*=.146, or the Read Aloud, *t*(21)=.193, *p*=.849.

APPENDIX F ADDITIONAL ANALYSES FOR EXPERIMENT 4

Mean proportions

The mean proportions of yes responses were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, *F*(2, 30)=49.358, *MSe*=7.629, *p*<.001, (η^2=.767). All contrasts were significant, *p's* <.001.

An alternative measure of sensitivity: d'

The *d'* scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, *F*(2, 30)=9.681, *MSe*=.144, *p*=.001, (η^2=.392). All contrasts were significant, *p*'s <.05.

Response bias

The B''_D scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant effect of condition, *F*(2, 30)=7.624, *MSe*=.030, *p*=.002, (η^2=.337). Contrasts showed that the B''_D scores were significantly lower in the Sing condition (*M*=.586, *SD*=.378) compared to the Read Aloud condition (*M*=.748, *SD*=.202; *t*(15)=2.195, *p*=.044) and compared to the Read Silently condition (*M*=.819, *SD*=.201; *t*(15)=3.601, *p*=.003. There was a marginally significant difference in the B''_D scores for the Read Aloud condition compared to the Read Silently condition, *t*(15)=1.779, *p*=.095.

APPENDIX G ADDITIONAL ANALYSES FOR EXPERIMENT 5

Mean proportions
The mean proportions of yes responses were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, *F*(2, 84)=67.327, *MSe*=17.477, *p*<.001, (η^2=.616). All contrasts were significant, *p's* <.05.

An alternative measure of sensitivity: d'
The *d'* scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant main effect of condition, *F*(2, 84)=36.923, *MSe*=.192, *p*<.001, (η^2=.468). All contrasts were significant, *p's* <.05.

Response bias
The B''_D scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a within-subjects factor. There was a significant effect of condition, *F*(2, 84)=5.814, *MSe*=.053, *p*=.004, (η^2=.122). Contrasts showed that the B''_D scores were significantly lower in the Sing condition (*M*=.549, *SD*=.380) and Read Aloud (*M*=.558, *SD*=.421) conditions compared to the Read Silently condition (*M*=.700, *SD*=.327; *t*(42)=3.586, *p*=.001 and *t*(42)=2.727, *p*=.009, respectively). There was no significant difference in B''_D scores for the Sing versus Read Aloud conditions, *t*(42)=.154, *p*=.878.

APPENDIX H ADDITIONAL ANALYSES FOR EXPERIMENT 6

Mean proportions

The mean proportions of yes responses were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a between-subjects factor. There was no significant main effect of condition, *F*(2, 57)=.330, *MSe*=291.262, *p*=.720, (η^2=.616). None of the contrasts were significant, *p's* >.435.

An alternative measure of sensitivity: d'

The *d'* scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a between-subjects factor. There was no significant main effect of condition, *F*(2, 57)=1.919, *MSe*=.217, *p*=.156. Although none of the contrasts were significant, *p's* >.05, it is important to note that *d'* scores were marginally greater in the Read Aloud (*M*=1.504, *SD*=.521) compared to the Read Silently condition, (*M*=1.212, *SD*=.435; *t*(38)=1.874, *p*=.069).

Response bias

The B''_D scores were analyzed in a one-way ANOVA with condition (Sing, Read Aloud, Read Silently) as a between-subjects factor. There was no significant effect of condition, *F*(2, 57)=2.278, *MSe*=.195, *p*=.112. Despite the non-significant main effect, planned contrasts were conducted. Contrasts showed that the B''_D scores were marginally greater in the Read Aloud condition (*M*=.636, *SD*=.285) compared to the Sing (*M*=.366, *SD*=.527; *t*(38)=2.019, *p*=.051) and Read Silently conditions (*M*=.392, *SD*=.476; *t*(38)=1.965, *p*=.057). The contrast for the Sing versus Read Aloud condition did not approach significance, p=.869.

APPENDIX I ADDITIONAL ANALYSES FOR COSTS AND BENEFITS

An alternative measure of costs and benefits: A' scores

We analyzed the *A'* scores for a within- and between-subjects experiment (Experiments 2 and 6, respectively) in a mixed repeated measures ANOVA with production condition (Sing, Read Aloud, Read Silently) as a within-subjects factor and study design (Within-Subjects, Between-Subjects) as a between-subjects factor. There was significant main effect of production condition, $F(2, 76)=9.834$, $MSe=.003$, $p<.01$, ($\eta^2=.206$), indicating a robust overall production effect for Sing and Read Aloud conditions. There was also a significant main effect of study design, $F(1,38)=7.090$, $MSe=.007$, $p=.011$, ($\eta^2=.157$), indicating overall greater *A'* scores in the Between-Subject study design ($M=.811$, $SD=.035$) compared to the Within-Subjects study design ($M=.770$, $SD=.035$). There was a significant two-way interaction between condition and study design, $F(2, 76)=6.949$, $MSe=.003$, $p<.01$, ($\eta^2=.155$), indicating a significant production effect for Sing and Read Aloud conditions in the within-subjects design and no significant production effect for Sing and Read Aloud conditions in the between-subjects design (see Experiments 2 and 6 for analyses). The presence of a production effect for Sing and Read Aloud conditions in the within-subject design reflected a significant cost to reading silently in the within-subject design compared to the between-subjects design, $t(38)= 2.997$, $p<.01$. There was no significant benefit to singing in a within-subjects design compared to a between-subjects design ($p>.450$); in fact, there was a significant cost associated with reading aloud in a within-subjects design ($M=.765$, $SD=.072$) compared to a between-subjects design ($M=.828$, $SD=.047$; $t(38)= 3.256$, $p<.01$). Taken together, these findings suggest that there does not seem to be a benefit for producing items via singing or reading aloud, but rather there seems to be a cost associated with reading items aloud and silently. Moreover, conducting a similar analysis to Bodner et al. (2014), we examined "benefits-over-silent" by comparing *A'* scores for the Sing and Read Aloud conditions in the within-subjects design to *A'* scores for the Read Silently condition in the between-subjects design. In contrast to both Bodner et al. (2014) and Forrin et al. (2016), there were no significant "benefits-over-silent" for the Sing, $t(19)=1.359$, $p=.190$ or Read Aloud conditions, $t(19)=1.332$, $p=.199$.