**Choosing what to read out loud while studying:**

**The role of agency in production**

by

Daniel Todorovic

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2020

**Examining Committee Membership**

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

| | |
|---|---|
| External Examiner | Steve Joordens<br>Professor, Department of Psychology, University of Toronto Scarborough |
| Supervisor | Colin M. MacLeod<br>Professor, Department of Psychology |
| Internal Members | Myra A. Fernandes<br>Professor, Department of Psychology |
| | Jonathan A. Fugelsang<br>Professor, Associate Chair of Graduate Affairs, Department of Psychology |
| Internal-external Member | John Turri<br>Professor, Canada Research Chair in Philosophy and Cognitive Science, Department of Philosophy |

**Author's Delcaration**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

**Abstract**

Both the acts of reading aloud (the production effect) and of making a choice are known to be beneficial for memory, however these benefits are yet to be established outside of typical laboratory settings. In this dissertation, I provide evidence to demonstrate the viability of production and choice for improving memory in the domain of academic studying. Experiments 1 and 2 show that a benefit of pure list production (in which the entire text is read aloud) can be obtained with lengthy narrative texts, while also showing that production benefits materials that have already been read before. Experiment 3, in which some participants chose the parts of a text to read aloud, demonstrates that production and choice can be used in tandem to benefit memory without any of the cost usually seen with production alone. Experiments 4 and 5 dissociated the effects of production and choice and demonstrated a boundary condition wherein one effect can overshadow another. Taken together, these findings show that production and choice – both being simple and powerful techniques – are viable and effective study strategies.

**Acknowledgments**

I would like to thank Colin MacLeod, my supervisor, for all the support, help, and advice that he's given me, as well as David McLean and David Lepinoy for their aid in carrying out the experiments.

I would also like to thank my family for their support in my endeavors.

**Table of Contents**

**List of Figures**

**List of Tables**

## Chapter 1

Throughout the world, people pursue higher education with the hope of finding a good job or just out of a desire to learn more. Of course, the process of achieving higher education requires a considerable amount of rigorous study. It is this requirement that concerns both students and educators, as students need to find ways to study effectively, and educators need ways to determine whether that studying has been effective—which is often assessed through tests of knowledge. Improving the effectiveness of study will pay off by enhancing learning and increasing student success. In this dissertation, I will explore two simple, powerful mnemonic strategies that may be used to aid memory for what is studied, and in everyday learning situations in general: the benefits of production and of choice. Before introducing these strategies, when considering how to aid memory while studying, it would be prudent to ask first how people typically study.

Of course, much of the required information in a given academic course can often be found in a textbook assigned to that course, along with whatever notes are taken by the student. It is common for courses to have a corresponding textbook from which required information can be learned, and it is expected that a student engages with these required readings while studying. Contrary to this expectation, Sikorski et al. (2002) found that introductory psychology students spent less than 3 hours a week reading their textbooks, preferring class notes. However, there is evidence that textbooks are engaged prior to testing. Berry, Cook, Hill, and Stevens (2010) found that among finance students, only approximately 40% read more than one hour a week, and that the greatest motivation to read was for exams, with the least endorsed reason for reading being to be prepared for class. In line with this, Clump, Bauer, and Bradley (2004) reported that among different psychology classes, students only read about 27% of the assigned texts before classes, but read about 70% of assigned readings in preparing for exams. Students also reported a high endorsement of study guide usage when studying, and indicated

that they would prefer that their professors inform them of what material is important for the tests (Berry et al., 2010), indicating a preference to differentiate important from less important information.

For a study strategy to be effective, it must first be practical, preferably simple, and should be used to aid memory of information believed to be important. Fortunately, a number of different ways to improve memory have been documented thus far. Mental imagery—imagining pictures of the information—has been shown to produce a considerable mnemonic advantage (Paivio, 1971). Enacting the actions indicated in a proposition has also been shown to result in improved memory for that proposition (Cohen, 1981). For example, the statement "Throw the ball" is better remembered when an individual subsequently throws a ball or watches someone else throw one than when the individual simply reads the phrase. Similarly, research on the generation effect has found that responses generated from cues are better remembered than those that are simply read (Slamecka & Graf, 1978). An example of a response generated from a cue might be "Toucan" generated from "Neotropical bird with a large, colourful beak."

Brown, Roediger, and McDaniel (2014) have reviewed these techniques and others in a book aimed at the everyday learner. All of these classic effects are well known, but they share a limitation: They are difficult to implement in daily activities. In studying, for example, not all study materials are easily imagined, such as abstract terms and concepts and complex propositions. The enactment effect may likewise be poorly suited to some study materials. Some studiers may also be embarrassed to perform actions for all study materials, especially if in public. The generation effect is a promising study aid but it requires the studier to first create separate cue materials based on the study materials, which may not be worth the effort for the benefit that doing so provides.

For a study strategy to be useful it must be effective and it must be simple to perform. The fewer limitations it has and the fewer demands it places on the learner, the better. One such simple,

powerful, and robust technique is the production effect. Put simply, a word read aloud is better remembered than a word read silently (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). This strategy bears a "family resemblance" to the other strategies just described, namely in the transformation of to-be-remembered items. Much like enactment, production involves performing an action on an item, but with production the action is much simpler to perform and does not suffer the same limitations concerning what can be acted upon.

The mechanisms underlying production are still being explored, but the distinctiveness account—that produced items are more memorable because they are made more more distinctive than non-produced items—has been the dominant explanation (Icht, Mama, & Algom, 2014; Lin & MacLeod, 2012; MacLeod et al., 2010; Ozubko & MacLeod, 2010; Ozubko, Major, & MacLeod, 2014; Richler, Palmeri, & Gauthier, 2013). The produced items are thought to be distinctive by contrast to non-produced items: Being spoken aloud makes them different from words read silently, and thus also makes them more memorable. This differs from a simple strength account, as distinctiveness need not involve the strengthening of a memory. Importantly, distinctiveness refers not to a property of the item but rather to the encoding process performed by the observer, such that distinctive information is processed along with the item (Hunt, 2006). In the case of production, as some items are spoken aloud, these items would be processed differently than items read silently. In particular, the aloud items would be encoded including a record that they were spoken, distinguishing them from the items read silently.

This distinctive record can in turn aid later memory for produced items, as they can easily be recognized as having been read aloud before. The use of distinctive information at test constitutes a "distinctiveness heuristic," which can also be used to reduce false alarms to lures during recognition testing (Dodson & Schacter, 2001), and this has been shown to apply in production (Forrin, Groot, & MacLeod, 2016). That the production effect is absent when the same non-unique response is given to

every item (MacLeod et al., 2010) and when production is applied to exemplars from a very limited number of categories (Richler et al., 2013) strongly implicates distinctiveness as the primary mechanism underlying the effect.

One aspect that the above cited studies have in common, aside from supporting a distinctiveness account, is that they all investigated the production effect with single items in lists. This is, in fact, standard procedure in the memory literature and is a result of designing experiments with emphasis on a high degree of experimental control. Of course, when involved in academic studying, people rarely study individual words in isolation. Only a very small number of studies have investigated the production effect with lengthier materials (Icht, Ben-David, & Mama, 2020; Ozubko, Gopie, & MacLeod, 2012) from which one might be able to extend production to a more realistic setting such as studying.

Finally, there is one important issue with respect to production and studying that, as yet, has not been investigated: the matter of choice. In most realistic settings that involve audible speaking, the speaker has the agency required to decide to speak aloud and also has the choice of what specifically to say aloud and what not to say aloud. This is also true in studying, of course. Except in situations of reading aloud without awareness of doing so, a studier intentionally chooses whether to read aloud and, if so, what portions of the material are to be read aloud. Given the ubiquity of academic studying, as well as the ease of producing, this venture is undoubtedly a very important one. In this dissertation, I will be examining the potential benefits of producing and of choosing on studying.

## The Production Effect

As already mentioned, the production effect is the finding that memory for produced words is superior to memory for non-produced words. Although named by MacLeod et al. (2010), the effect had been noted or hinted at by researchers for at least a century. Quantz (1897, p. 36) noted that “motor-

minded" individuals—that is, individuals who make lip movements when reading—may be aided in comprehension of written materials, if only because of the extra time it takes to produce such lip movements, keeping them focused longer on the material being read. Gates (1917), in an early work on the mnemonic effects of recitation, noted that his participants reported that production was an effective aid for retrieval. Bartlett (1932) noted that individuals who vocalized while learning drawings of faces were less likely to attribute false details to the faces when later describing them than were individuals who did not vocalize (p. 58), suggesting that vocalizers may be more cautious in their mental activity (p. 112).

Barlow (1928) may have been the first to directly test the effects of reading aloud versus silently on memory. Participants first read two lists of consonant-vowel-consonant non-words (CVCs such as "daf"), one list out loud and the other list silently. On a subsequent free recall test, performance was greater for the list that had been read aloud. But it was almost half a century before this phenomenon became the subject of renewed study. The current era of production effect research began with Hopkins and Edwards (1972), who observed that recognition of words read aloud earlier was better than recognition of words read silently earlier. They also observed that this was true only for the within-subject, mixed list situation where participants read some words aloud and some words silently; in their study, there was no observable difference in recognition in a between-subjects, pure list situation, where one group read all of the words aloud and another group read all of the word silently.

Since the Hopkins and Edwards (1972) study, the production effect has been replicated many times and has been shown to be a robust effect (e.g., Conway & Gathercole, 1987; Forrin & MacLeod, 2018; Hassall, Quinlan, Turk, Taylor, & Krigolson, 2016; Icht et al., 2014; MacDonald & MacLeod, 1998; and so on). MacLeod and Bodner (2017) provide a summary of this research. Production need not involve speaking as long as something is produced, such that production effects have been found with

writing and typing (Forrin, MacLeod, & Ozubko, 2012; Jamieson & Spear, 2014; Sundqvist, Mäntylä, & Jönsson, 2017), singing (Hassall et al., 2016; Quinlan & Taylor, 2013, 2019), and even mouthing words silently (Castel, Rhodes, & Friedman, 2013; Conway & Gathercole, 1987; Forrin et al., 2012; MacLeod et al., 2010). A mnemonic benefit is also present in the absence of self-production, as when hearing someone else produce a word compared to reading it silently (Forrin & MacLeod, 2018; Gathercole & Conway, 1988; MacLeod, 2011), and even when imagining the typing of a word (Jamieson & Spear, 2014). Production may even help prevent forgetting, as produced words are less susceptible to directed forgetting (Hourihan & MacLeod, 2008).

As already mentioned, production is much less robust in pure list situations. When the Hopkins and Edwards (1972) finding of no production effect with pure lists was replicated by MacLeod et al. (2010), and given a similar finding by Dodson and Schacter (2001), the belief was that there was no pure list production effect. Since then, Fawcett (2013) and Bodner, Taikh, and Fawcett (2014) have demonstrated that there is a pure list production effect, replicating the results of Gathercole and Conway (1988), albeit observing a considerably smaller effect than is seen in mixed list situations. This finding of a pure list production effect lends credence to a strength-based contribution to the effect—that the aloud items have greater memory strength than the silent items (see Bodner & Taikh, 2012; Jamieson, Mewhort, & Hockley, 2016). In a pure list situation, no item would be processed distinctly from another on the basis of production, so differential strength would explain the observed effect; in a mixed list situation, both strength and distinctiveness could be operating, thereby producing a larger effect.

Bodner, Jamieson, Cormack, McDonald, and Bernstein (2016) found that when the duration of silent items was lengthened during study, the pure list production effect vanished whereas the mixed list effect remained. This suggests that for a mixed list, where distinctive information is more available,

distinctive processing is the better retrieval cue, which is not the case in pure lists, where memory strength is thought to be more useful. However, as the silent items are made more memorable in such a way—such that one would expect performance for the pure silent group to increase—and that pure list effects are generally small, one cannot rule out distinctiveness for pure list production effects. Rather, it may simply be that silent items are being strengthened to the level of the more distinctive aloud items. Indeed, the assumption that distinctiveness would not aid in pure lists is not beyond rebuttal. In recognition tests, at least, any non-studied lure/distractor is, by default, a silent item and can provide contrast for the studied items, all of which were produced. Some participants may not think to employ such a distinctiveness heuristic, as distinctiveness would not have been apparent during study. Some participants may use the strategy only sporadically throughout the test. And some may not realize that this strategy is useful until part way through the test. This variation in use of the distinctiveness strategy could account for the smaller pure list than mixed list production effect. At any rate, the important finding is that production works in both types of lists, with more robust effects in a mixed list.

The production effect, although demonstrated to be an effective memory aid, is not without its drawbacks. For the considerable benefit it confers, the production effect also has an associated cost. In recognition testing, as first found by Hopkins and Edwards (1972), there is a cost to silent items in mixed lists such that they are more poorly remembered compared to items in a pure silent list. This cost has been replicated multiple times since then (Bodner et al., 2014; Forrin et al., 2016). As to why this cost occurs, the most intuitive explanation is a “lazy reading” account—originally proposed by Begg and Snider (1987) for a similar cost seen in the generation effect—which suggests that participants decide that produced items are in some way more important than non-produced items, resulting in less attention to and processing of non-produced items in mixed lists.

Alternatively, it is possible that this cost to silent items is a feature of distinctiveness: During recognition testing, should a participant use distinctiveness as a retrieval heuristic, this would certainly aid in correct rejections of new items but at the same time might also result in a cost to silent items. This is because both silent items and new items have not been produced. During recognition testing, a participant might try to determine whether an item had been previously read aloud and, deciding that it had not been, might reject that item as being a non-studied item despite it having been previously studied silently (Forrin et al., 2016; Ozubko, Bamburoski, Carlin, & Fawcett, 2020). Needless to say, in pure list recognition testing, there is no cost to silent items.

Turning to recall testing, the production effect in mixed lists seems to be driven largely by costs and there appears to be no production effect in pure lists (Jones & Pyc, 2014; Lambert, Bodner, & Taikh, 2016; Forrin & MacLeod, 2016). This mixed list cost persists even after generation and imagery tasks, which casts some doubt on the lazy reading hypothesis because of the more elaborative encoding involved in these encoding conditions (Lambert et al., 2016). This cost-driven production effect in recall might be best explained by a disruption of item order. In mixed lists, reading some words aloud and others silently might result in enhanced elaboration for produced items, but at the expense of disrupting memory for item order (Jonker, Levene, & MacLeod, 2014; Jonker & MacLeod, 2015). The item order account was proposed by McDaniel and Bugg (2008) as an explanation for why several encoding techniques that benefit free recall in mixed lists do not produce corresponding benefits in pure lists. Their central hypothesis is that, in mixed lists, unique or different items disrupt encoding of order information of other items while recruiting greater amounts of item-specific encoding for themselves. In the context of production, produced words—being unusual compared to silently read words—benefit from greater item-specific encoding, whereas the presence of multiple unusual, produced words in the list disrupts order information for the less unusual silently read words. Pure list scenarios, in which all items are processed the same way, would encourage the use of item order information as a retrieval

strategy. As is now obvious, there are a number of different explanations of the cost to silent items in the production effect literature.

As mentioned earlier, the vast majority of production effect research has used lists of single items. Some researchers have, however, ventured into paired associate production (Hourihan & Smith, 2016; Krishnan, Watkins, & Bishop, 2017; Ozubko et al., 2012; Putnam, Ozubko, MacLeod, & Roediger, 2014), a few have even investigated the production effect with full sentences (Ozubko et al., 2012; Icht et al., 2020), and there was one study that had participants read the last word in a sentence aloud or silently (Rommers, Dell, & Benjamin, 2020). Concerning reading full length articles, similar to texts one would expect to read over while studying, investigations of the production effect are scarce (Ozubko et al., 2012; Kline, 2019). The data that do exist concerning production while studying test materials are not completely informative as to the usefulness of production, and there remains a considerable amount of uncertainty.

Ozubko et al. (2012) first investigated production with full texts. In their study, participants read an article, similar to what one might find in a textbook, by reading some paragraphs aloud and others silently. Participants then had their memory tested by fill-in-the-blanks tests, either immediately after reading or after a one-day retention interval. At both testing intervals, a production effect was obtained, demonstrating the benefit of producing while studying as well as showing that this benefit can endure longer intervals. Importantly, Ozubko et al. (2012) used a mixed list design but did not attempt a pure list design. Kline (2019) attempted to replicate this production effect with longer texts but with a pure list design instead of a mixed list design, but failed to find a significant difference between aloud and silent groups. Kline reported that her sample size could have been too low to find an effect (around 20-30 per group). This certainly seems possible, especially given that pure list production effects are generally less robust than mixed list effects, and full length articles present a largely unexplored territory

for production. As such, it is too early to say whether a pure list production effect can be obtained with full-length articles. This is one of the questions that motivated this dissertation.

A key issue that neither Ozubko et al. (2012) nor anyone else has yet investigated is the effect of choosing while producing. Ordinarily, when a person is studying, if they decide to study aloud, the portions of the material that they choose to study aloud are up to them. Does choosing what to produce matter? Does the choice itself influence retention? Without knowing the answer to these questions, the generalizability and usefulness of production while studying cannot be fully established. Choosing is not just an action without consequence: It can have effects of its own. To understand how choice might affect memory in tandem with production, it is necessary to turn to the literature on choice during study.

## The Choice Effect

As it turns out, another simple yet effective strategy to aid remembering is to choose one option from a selection of available options: Memory for the chosen option is aided by having made the choice. The choice effect, or self-choice effect as it was called by Takahashi (1989; as cited in Watanabe, 2001), has been shown to significantly influence memory (Perlmuter, Monty, & Kimble, 1971; Watanabe, 2001; and others). Although much less widely known, a small and continuing literature exists on the effects of choice on memory. The concept itself is simple: The participant is given a number of options from which to choose. The criterion that they use to choose depends on the study in question, but the important feature is that a choice is made. The typical findings are that memory for the chosen option is better than memory for the alternatives and that memory for the chosen option is better when they choose it themselves than when that option is selected for them by the experimenter or another person (Cloutier & Macrae, 2008; Coverdale & Nairne, 2019; Monty, Perlmuter, Libon, & Bennet, 1982; Murty, DuBrow,

& Davachi, 2015; Perlmuter et al., 1971; Takahashi, 1992; Toyota & Tsujimura, 2000; Watanabe, 2001; and others).

This literature began with Perlmuter et al. (1971), who examined the effects of choice in paired-associate learning. Their general finding was that, when participants were asked to choose which response was to be learned in an A-B pair, if they chose B, for example, then learning of B was facilitated and also resulted in more interference in subsequent A-C learning. This benefit in paired-associate learning has been replicated a number of times since (Monty, Geller, Savage, & Perlmuter, 1979; Monty & Perlmuter, 1972; Monty, Rosenberger, & Perlmuter, 1973; Perlmuter, Scharff, Karsh, & Monty, 1980).

Of course, the benefit of choice has been extended to recognition (DuBrow, Eberts, & Murty, 2019; Monty et al., 1982; Watanabe, 2001; Driver, 2013) and to recall (Hirano & Ukita, 2003; Watanabe & Soraci, 2004; Toyota, 2016; Coverdale & Nairne, 2019). In recognition, memory for the chosen item is better than memory for the non-chosen alternatives; the chosen item is also better remembered than a pre-selected item under forced-choice conditions (i.e., when the chosen item was selected by someone else). Of considerable interest is the robust finding that it is not just the chosen item that benefits: Memory for non-chosen alternatives is also improved in choice conditions compared to forced-choice conditions (Monty et al., 1982; Perlmuter & Monty, 1982; Watanabe, 2001; Watanabe & Soraci, 2004). For example, when choosing option A between options A, B, and C, not only is memory for A improved but so is memory for B and C, although not as much as A. Why this might be the case is easy to imagine. When a participant is deliberating among a number of alternatives, they likely engage with all of the options more than if one option is chosen for them by another person. It has been suggested that the benefit to the chosen item might be due to distinctiveness, as the chosen item has the distinction of being, well, chosen (Takahashi, 1997; as cited in Watanabe & Soraci, 2004). In free recall, however, the effect for non-chosen options is not clear: A number of studies have shown the benefit of choice to be

limited to chosen items (Toyota, 2015; 2016) but recall of non-chosen items in these studies was at floor, so it is possible that under different circumstances an effect of choice might be found even for the non-chosen options. This interpretation is supported by the findings of Hirano and Ukita (2003), who provided evidence that recall of non-chosen options can benefit from choice.

Whereas the benefit and cost of production are straightforward, whether there is a cost associated with choice has not been determined: Nearly all studies have investigated choice in a within-subject design. In one of the only between-subjects choice studies in the literature, Monty et al. (1982) investigated choice effects on recognition and showed a benefit to chosen items and to non-chosen alternatives. Thinking about the matter rationally, there is little reason to assume the existence of a cost given that non-chosen items generally show a benefit in recognition, in contrast to non-produced items in production effect research.

It is possible that pre-selected targets in forced-choice trials incur a cost when pre-selected trials are intermixed with choice trials. The few studies to directly compare within-subject to between-subjects designs in regard to choice do not support this possibility. Monty and Perlmuter (1972) and Monty et al. (1973) both found that, contrary to costs, when choice and pre-selected trials were mixed, learning of the pre-selected response in paired-associate learning was also facilitated, compared to a pure list pre-selected group. These results used the paired associate learning paradigm but no similar comparison has been done with recognition or recall. However, from these results, we can expect not to see a cost and maybe even to see a benefit to pre-selected targets in recognition or recall should a direct comparison be made.

As may be apparent by now, the choice literature is limited in the number of studies investigating between-subjects manipulation of choice and no-choice conditions. [Henceforth, I will refer to the no-choice condition as the “force” condition, for greater clarity.] Although memory for

chosen items has been shown to be superior to memory for non-chosen alternatives and forced items, and memory for non-chosen alternatives receives a benefit as well when deliberated over, it is still difficult to say what effect choice trials have on force trials in a mixed design without an appropriate baseline for comparison. This line of research would be of great interest in the future, but goes beyond the scope of this dissertation, which is concerned with production and the possible interaction of production with choice.

Although the mechanism likely underlying production is distinctiveness, the mechanism or mechanisms underlying the choice effect is more uncertain. The early research on choice proposed a motivational account where having agency raises participant motivation, hence improving memory (Monty et al., 1973; Perlmuter et al., 1980). Of course, this just pushes the problem back a level: What does this greater motivation lead people to do differently in terms of learning and memory processes? Alternatively there is a metamemory account, wherein participants strategically choose items on the basis of how easy they believe those items will be to remember (Takahashi, 1992); this account has some support from the self-directed learning literature (Kornell & Metcalfe, 2006; Metcalfe, 2009). The multiple-cue hypothesis, proposed by Soraci et al. (1994) to offer an explanation for the generation effect, has also been proposed as an explanatory account of choice wherein participants generate multiple retrieval cues while making a choice (Watanabe, 2001; Watanabe & Soraci, 2004). Several other accounts exist as well, although the multiple-cue hypothesis seems very straightforward and plausible. However, resolving the issue of what mechanisms drive the effect of choice is beyond the scope of this dissertation.

As mentioned, there is a possibility that selecting an item may make it more distinct compared to non-chosen alternatives, which is a separate phenomenon from the benefit seen for non-chosen alternatives when choosing. Indeed, even when participants cannot deliberate on to-be-remembered

items because they are unaware of the options presented to them, the simple act of choosing an option blindly can result in a benefit to the chosen items (Cloutier & Macrae, 2008; Cunningham, Brady–Van den Bos, & Turk, 2011; Voss, Gonsalves, Federmeier, Tranel, & Cohen, 2011). It seems, then, that choice involves at least two different processes, and Driver (2013) accordingly proposed a two-stage model of choice. The first stage is what he called deliberation, in which the observer contemplates the available options. The second stage is selection, in which the actual choice is made. One may reasonably assume that the observed benefit for non-chosen items under choice conditions is due to deliberation, whereas the chosen item itself can benefit from both deliberation and selection.

From this analysis, it seems likely that there are different mechanisms underlying production and choice, rather than the two strategies recruiting the same mechanisms. As such, it is reasonable to suppose that there may be a benefit in studying when using each of these techniques, and a critical question becomes how these two encoding operations work in concert. For the purposes of this dissertation, Driver's (2013) two-stage process will be assumed.

## Rationale for the Experiments

To date, with the exception of the Ozubko et al. (2012) and Kline (2019) studies, production effect research has primarily used lists of unrelated items, some produced and some not produced during the study phase. A robust advantage for those that have been produced has been uniformly observed, at least when the two types of encoding—aloud and silent—are mixed. Is this simple technique valuable to apply in the "real world"? The Ozubko et al. study, using extended texts, suggests that it may be, but they examined only the within-subject, mixed list situation. They also had participants read a text only once, and what they read aloud was specified for them by the experimenter. These constraints limit the generalizability of the Ozubko et al. findings in terms of the broader realm of studying textual materials. The purpose of this dissertation is to remove these

restrictions and to examine whether production may benefit learning and memory in situations where texts are read entirely aloud, where they are read more than once, and ultimately where the reader chooses what to read aloud. Table 1 provides a sketch of the series of experiments.

The dissertation begins with a simple experiment in which participants read a whole four-page novel text once only, either all aloud or all silently—the between-subjects, pure list situation. Does production confer an advantage in this situation, as it did when participants read parts of a text aloud in the Ozubko et al. (2012) study? Moving on from there, it is most likely that people ordinarily read texts multiple times, as when they read texts when they are assigned during the term and then again later to review the material when it is time for them to be tested. Consequently, Experiment 2 had participants read texts twice, either with both readings done silently or with one of the readings done entirely aloud. Experiment 2 also served as a "bridge" to the central experiment in the dissertation—Experiment 3.

**Table 1.** Overview of the Experiments

| Study | Purpose of Investigation | Materials | Manipulation |
|---|---|---|---|
| Exp. 1 | Pure list production; one reading | Toucan article | Aloud vs Silent (between-subjects) |
| Exp. 2 | Pure list production; two readings | Toucan article | Aloud vs Silent (between-subjects) |
| Exp. 3 | Production and choice; two readings | Toucan article | Aloud vs Silent (within-subject), Choice (between-subjects) |
| Exp. 4 | Production-choice interaction | Sentence pairs | Aloud vs Silent (between-subjects), Choice (between-subjects) |
| Exp. 5 | Production-choice interaction | Sentence pairs | Aloud vs Silent (within-subject), Choice (between-subjects) |

In Experiment 3, all participants read the text silently once before any manipulation took place. This allowed them to "get the lay of the land," such that they had a good idea of the content. Then,

participants were divided into three groups. In the Control group, on the second reading, they again read the entire text silently. In the Choice group, they were instructed to choose about a third to a half of the text to read aloud, with the goal of anticipating and emphasizing what they would expect to be tested on. Each participant in the third group, called the Force group, was yoked to a participant in the Choice group, such that underlining in the text indicated what was to be read aloud and corresponded to what the yoked participant in the Choice group had actually chosen to read aloud. This experiment was designed to capture the essence of what studiers do ordinarily: After having read the materials once, when they return to it, they choose what to emphasize. Does choice in terms of what to emphasize confer an advantage, in addition to any advantage conferred by production?

Experiments 4 and 5 pursued the potential interaction of production and choice, two encoding manipulations that are each known to produce benefits in standard list-learning experiments (as summarized in the introduction to this dissertation). Can we separate the contributions of choice and production? Can we determine the extent to which they interact? Put simply, if production provides an advantage when studying text, does choosing what to produce increment that advantage?

This dissertation has the empirical goal of extending production beyond laboratory learning of unrelated and isolated materials to more meaningful, extensive materials. It has the practical goal of determining whether production—and choice—can be useful in real-world studying of meaningful material. And it has the theoretical goal of contributing to our understanding of the mechanism(s) underlying the production effect, notably those of distinctiveness and strengthening, by considering why production helps remembering.

## The Experiments

The first question motivating the experiments in this dissertation was whether the within-subject production benefit observed by Ozubko et al. (2012) when participants read parts of a scholarly

article out loud and other parts silently would extend to the between-subjects situation. Would participants who read an entire article out loud remember its content better than participants who read the entire article silently? Should this be the case, then production would be among the simplest study techniques to use to aid academic performance. Experiments 1 and 2 in particular address this question.

The primary goal of the dissertation was to determine the influences of production and choice during study of text on later memory for that text. Ordinarily, people choose what they want to read aloud, so Experiments 3 through 5 were designed to address this goal. As described in the introduction, there is reason to believe that both production and choice should be beneficial to studying, but it is uncertain how they would differentially influence retention of study materials—and whether and how they might interact. The principal prediction was that, when reading some of the content out loud, participants given choice as to what to read aloud would remember those portions of the text better than would participants who read identical materials out loud but without having chosen them—who would in turn remember better than participants who only read the text silently. To enable this comparison would require having the no choice group yoked to the choice group, such that the two groups read the same material aloud. Thus, what a given participant chose to read aloud in the choice group would be assigned to a yoked participant in the no choice (Force) group.

In Experiments 3 through 5, several possible patterns could be observed, all anticipating a benefit of production. First, the memory benefit for produced items could be even greater when participants choose what to produce—that is, the mnemonic benefits of production and choice could be additive. Second, memory for both produced and non-produced items could be enhanced when participants are able to choose what to produce. And third, memory for unproduced items could be greater when participants choose what to produce: Although production is associated with costs to non-produced items (Jonker et al., 2014; Forrin et al., 2016), choice is associated with benefits to non-chosen

items compared to when participants must ”choose” an item pre-selected by the experimenter (Perlmuter & Monty, 1982; Watanabe, 2001), as choosing between alternatives requires evaluating options that were not chosen. Of course, it is possible that the benefits of production and choice are mutually exclusive: There is some reason to expect that when some encoding or retrieval strategies are combined, then participants may forego one strategy in favour of another (see Putnam et al., 2014; see also the General Discussion for more on the possibility of “overshadowing”).

Experiment 1 was a first step, designed to extend the production effect to studying academic materials all aloud versus all silently. Participants would read an approximately 2000-word article, either out loud or silently, and then have their memory tested via a fill-in-the-blanks test, a quite common test format. Experiment 2 was designed to bridge Experiments 1 and 3, as Experiment 3 would use a procedure—rereading—that was novel to production effect research. Thus, Experiment 2 was conducted to introduce the procedure and provide a link between the two experiments. Participants in Experiment 2 were required to read the article twice, both times silently or once silently and once aloud, to determine how having knowledge of the content before invoking production would influence the production benefit. Experiment 3 was the critical experiment, in which participants again read the article twice, as in Experiment 2, but on the second reading they could either choose what portions to read out loud, read pre-selected portions out loud, or read the entire article silently for a second time. Experiment 4 was a follow-up to Experiment 3, using simpler materials with the aim of disentangling the effects of production and choice. Finally, Experiment 5 was a conceptual replication of Experiment 4 designed to address a minor methodological issue. More rationale for each experiment will be presented in the introductions to the individual experiments.

## Chapter 2: Pure List Production while Studying

### Experiment 1A

Ozubko et al. (2012) demonstrated a production effect for text: When some paragraphs of an academic article were read aloud and some were read silently, memory was better for the material in those paragraphs that had been read aloud. This is effectively an extension of the mixed list design prevalent in production effect research. However, no one has yet investigated whether the production effect extends to longer educational articles when the entire article is read out loud versus silently—corresponding to a pure list design. If reading an entire article aloud is beneficial, this would then constitute a very simple study strategy with broad implications for school or work performance. The production effect is well established to be less robust in between-subjects designs, so it is possible that any potential benefits observed using the mixed list design would not extend to the pure list design. Experiment 1 addressed this question.

#### Method

**Participants.** 55 undergraduate students from the University of Waterloo (45 female, mean age = 19.81, *SD* = 3.14) participated for course credit. From the Aloud group, one participant was excluded due to whispering the text material rather than reading out loud; one participant was excluded due to rereading several sections of the text; one participant was excluded due to skipping several passages of the text while reading; and one participant was excluded for reading silently. From the Silent group, two participants were excluded for reading the text too fast (cutoff of 4.5 minutes determined by histogram analysis); and one participant was excluded for engaging in text messages during reading. This left 48 participants, 24 in each of the Aloud and Silent groups.

**Materials.** The materials for this experiment consisted of the Toucan article and the test sheet used by Ozubko et al. (2012),[1] which were in turn obtained from Chan, McDermott, and Roediger (2006). The Toucan article is approximately 1900 words in length, and was divided into 4 pages for this experiment. The article was presented to participants using E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). The test consisted of 24 fill-in-the-blanks questions based on the content of the article, and was administered on a printed test sheet. Performance during reading was recorded using Audacity® recording and editing software[2], which was used to determine compliance with task instructions. Both the article and the test questions—together with the correct answers—are presented in the Appendix.

**Procedure.** Participants were tested individually. They were asked to turn off their phones so as not to be distracted during the study. Participants sat at a computer and were told that they were going to read an article on which they would subsequently be tested. The four pages of the article were presented one page at a time on the computer monitor; participants were asked to press the spacebar to proceed to the next page. Participants in both the Aloud and Silent conditions were asked to read the entire article, without rereading or skipping sections. Additionally, participants in the Aloud condition were instructed to read the entire article out loud, without mumbling, as if they were reading to someone else. Participants in the Silent condition were asked to read the entire article silently, without whispering or moving their lips. Reading was self-paced.

[1] In a study not included here, participants read two of the Chan et al. articles—Toucan and Hong Kong—one Aloud and the other Silent. While the Toucan article was conducive to the production effect, the Hong Kong article was not, a pattern that Kathleen Hourihan had also observed in the Ozubko et al. (2012) study (personal communication, February 15, 2019). It may be the case that the production effect is limited when the reading material contains multiple unfamiliar words that are difficult to pronounce, as is the case with the Hong Kong article. This is consistent with the findings of Dahlen and Caldwell-Harris (2013) and Kaushanskaya and Yoo (2011).

[2] Audacity® software is copyright © 1999-2019 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.

After they had read the entire article, the experimenter provided participants with the test sheet and gave them as much time as they needed to complete it. Participants were encouraged to answer all of the questions, leaving none blank. Scoring of the test sheets was initially done using two different criteria: stringent and lenient. Under stringent scoring, an answer was considered correct if the participant could not have arrived at that answer in any other way than through having read that section of the article. For example, if the answer was "hemochromatosis" and the participant answered with "hemotosis," it would be scored as a correct answer. Under lenient scoring, an answer would be scored as correct if it was similar to the actual answer, even if it was wrong. For example, if the actual answer was "fly to the spirit world" and the participant answered with "fly in the sky," this would be accepted as correct. Stringent and lenient scoring produced identical patterns (with lenient scoring only resulting in slighter greater overall test performance), so the results will be reported only for the stringent scoring.

**Results and Discussion**

*Reading Times*

In terms of reading times during the study phase, as expected, the Aloud group ($M$ = 805.25 s, $SD$ = 124.60) read the article significantly slower than the Silent group ($M$ = 652.60 s, $SD$ = 167.45), $t(46) = 3.58$, $p = .001$.

*Memory Test: Fill-in-the-blanks*

Table 2 presents the means and standard errors for performance on the fill-in-the-blanks test for the two conditions for Experiments 1 and 2; Experiment 1A is in the first row. An independent samples $t$-test was conducted for memory performance on the fill-in-the-blanks test. Although performance for the Aloud group was numerically greater than performance for the Silent group, the groups did not differ significantly, $t(46) = .94$, $p = .36$, $d = .27$, $1\text{-}\beta = .15$. From the Aloud group, 15

participants had numerically higher scores than the Silent group average, in the direction of the production effect. However, 14 from the Silent group also showed this pattern.

These results might indicate that reading an entire article out loud does not result in a production effect, unlike when only parts of it are read aloud, as in the Ozubko et al. (2012) study. However, it is also a fact that the production effect is much smaller in pure lists than in mixed lists, and as a result sometimes is obtained (Gathercole & Conway, 1988; Fawcett, 2013; Bodner et al., 2014) and sometimes is not obtained (Hopkins & Edwards, 1972; MacLeod et al., 2010). It is possible that a small benefit may exist, but simply has not been observed here. For this purpose, Experiment 1B was conducted.

## Experiment 1B

Experiment 1B was conducted to determine whether a between-subjects production effect could be found by modifying the procedure of Experiment 1A. Experiment 1B was identical to Experiment 1A in all respects, except that Experiment 1B had a larger sample size (n = 30 per group), and the instructions were reworded in an effort to be more direct and clear (not because participants misunderstood the instructions, but merely in an attempt to enhance the likelihood of obtaining a production effect).

### Method

**Participants.** 76 undergraduate students from the University of Waterloo (55 female, mean age = 19.45, *SD* = 1.98) participated for course credit. From the Aloud group, data from seven participants were excluded due to skipping sections of the text; one participant was excluded for reading silently; one participant was excluded for rereading portions of the text; and the data from one participant were excluded because she reported brain damage that adversely affected her memory. From the Silent

group, data from four participants were excluded due to audibly reading the text; one participant was excluded for reading the text too fast (cutoff of 4.5 minutes determined by histogram analysis); and data from one participant were excluded because she had a phone call conversation during reading. This left 60 participants, 30 in each of the Aloud and Silent groups.

**Materials.** The materials in Experiment 1B were identical to those in Experiment 1A.

**Procedure.** The procedure of Experiment 1B was identical to Experiment 1A, except that the instructions were reworded to emphasize the production task (read out loud or silently), by reiterating the production task at the end of the instructions. As in Experiment 1A, stringent and lenient scoring produced identical patterns of results, with lenient scoring only producing generally higher performance, so only stringent scoring results will be reported.

**Results and Discussion**

*Reading Times*

In terms of reading times during the study phase, as expected, the Aloud group ($M$ = 786.33 s, $SD$ = 130.21) read the article significantly slower than the Silent group ($M$ = 625.50 s, $SD$ = 189.46), $t(58) = 3.83$, $p < .001$. This difference (161 s) was almost identical to that in Experiment 1A (152 s).

*Memory Test: Fill-in-the-blanks*

The memory performance data for Experiment 1B are presented in the second row of Table 2. An independent samples $t$-test was conducted for memory performance on the fill-in-the-blanks test. Although performance for the Aloud group again was numerically greater than performance for the Silent group, the groups did not differ significantly, $t(58) = .94$, $p = .35$, $d = .24$, $1-\beta = .15$. The number of participants from the Aloud group who performed greater than the Silent average was 17, compared to 15 from the Silent group who performed greater than the Silent average.

**Table 2.** Proportion of Correct Responses as a Function of Condition in Experiments 1 and 2

| | Aloud | Silent |
|---|---|---|
| Experiment 1A | .52 (.04) | .48 (.03) |
| Experiment 1B | .59 (.02) | .55 (.03) |
| Experiment 2A | .68 (.03) | .54 (.03) |
| Experiment 2B | .66 (.03) | .63 (.03) |

Note: Standard errors are shown in parentheses.

These results replicate the findings of Experiment 1A, implying that pure list production does not aid learning with long texts. Performance in Experiment 1B was overall a little higher than that in Experiment 1A, although it is not clear whether this is due to the reworded instructions or merely to sample differences. A similar increase was observed in Experiment 2B, and will be discussed later.

*Combined Analysis: Experiments 1A and 1B*

To further evaluate this conclusion of no difference, the data from Experiments 1A and 1B were combined, given their similarity, to gain more power. Here, a 2 x 2 analysis of variance (ANOVA) with Experiment (1A or 1B) and Encoding (Aloud or Silent) as between-subjects factors was conducted. An effect of Experiment was found, $F(1, 104) = 4.38$, $MSE = .03$, $p < .05$, $\eta_p^2 = .04$, $1-\beta = .55$, reflecting that participants in Experiment 1B performed better than those in Experiment 1A. However, neither the effect of Encoding, $F(1, 104) = 1.79$, $MSE = .03$, $p = .18$, $\eta_p^2 = .02$, $1-\beta = .26$, nor the interaction of Experiment and Encoding, $F(1, 104) = .03$, $MSE = .03$, $p = .86$, $\eta_p^2 = 0$, $1-\beta = .05$, was significant. This pattern of results was seen in Kline's (2019) research with pure list production using long texts as well. Kline suggested that the reason she failed to find an effect might have been due to her small sample size, but the results presented here suggest that pure list production may simply not be a beneficial strategy for improving memory while studying long texts, at least in reading them for the first time.

## Experiment 2A

In Experiment 2A, participants again read the Toucan article, except that this time they read it twice. For all participants, the first reading was always done silently; on the second reading, participants read the article either entirely aloud or entirely silently. There were two reasons for this procedural change. The first had to do with this experiment serving as a "bridge" to Experiment 3. When choice is introduced in Experiment 3, it would not be realistic for participants to try to choose text to read aloud without first knowing what they were choosing, so an initial silent reading was seen as the best way to mitigate this problem. Second, this also offered the first opportunity to observe whether a production effect could be obtained for information that has previously been read silently, which likely closer approximates what people would do normally in studying.

### Method

**Participants**. 55 undergraduate students from the University of Waterloo (45 female, mean age = 19.58, *SD* = 1.73) participated for course credit. From the Aloud group, data from one participant were excluded because the sound recording device failed to operate; one more participant was excluded due to skipping parts of the text. From the Silent group, data from one participant were excluded due to falling asleep during the reading phase; one participant was excluded due to repeatedly engaging in text messaging during the reading phase; and data from three participants were excluded for reading the text too fast (cutoff of 4.5 minutes determined by histogram analysis). This left 48 participants, 24 in each of the Aloud and Silent groups.

**Materials.** The materials in Experiment 2A were identical to those in Experiments 1A and 1B.

**Procedure.** The procedure in Experiment 2A was identical to that of Experiment 1A, except that all participants read the article twice, with the first reading being a silent reading. Participants in the

Silent condition thus read the article twice silently whereas participants in the Aloud condition read the article once silently followed by once aloud. As in Experiment 1, both stringent and lenient scoring procedures were conducted. Again, the pattern of data did not differ (with the lenient scoring resulting only in overall greater test performance), so the results are reported for the stringent scoring only.

**Results and Discussion**

*Reading Times*

In terms of reading times during the study phase, on the first reading (done silently for all participants), as expected, there was no group difference: The Aloud group ($M$ = 566.96 s, $SD$ = 153.19) took the same amount of time to read the article as did the Silent group ($M$ = 513.82 s, $SD$ = 141.49), $t(46) = 1.25$, $p = .22$. On the second reading, as in Experiments 1A and 1B, the Aloud group ($M$ = 770.50 s, $SD$ = 117.72) read the article significantly slower than the Silent group ($M$ = 488.75 s, $SD$ = 149.53), $t(46) = 7.25$, $p < .001$.

*Memory Test: Fill-in-the-blanks*

The memory performance data for Experiment 2A are presented in the third row of Table 2. An independent samples $t$-test was conducted for memory performance on the fill-in-the-blanks test. Performance for the Aloud group was significantly better than performance for the Silent group, $t(46) = 3.17$, $p < .01$, $d = .91$, $1-\beta = .87$. The Aloud group had 17 participants that performed numerically greater than the Silent average, compared to 10 from the Silent group who performed greater than the Silent average.

*Combined Analysis: Experiments 1A, 1B, and 2A*

These results clearly differ from those observed in Experiment 1, in which pure list production did not result in a significant production effect. To reinforce this conclusion, a 2 x 2 ANOVA was

conducted with Experiment (1A, 2A) and Encoding (Aloud, Silent) as between-subjects factors. Both the main effect of Experiment, $F(1, 92) = 11.06$, $MSE = .03$, $p = .001$, $\eta_p^2 = .11$, $1\text{-}\beta = .91$, and the main effect of Encoding, $F(1, 92) = 7.66$, $MSE = .03$, $p < .01$, $\eta_p^2 = .08$, $1\text{-}\beta = .78$, were significant, but there was no significant interaction, $F(1, 92) = .18$, $MSE = .03$, $p = .18$. This is likely because the Aloud group in Experiment 1A performed numerically better than the Silent group, so the difference was in the same direction in both experiments, despite not being significant in Experiment 1A (or in Experiment 1B). Similarly to the above, another 2 x 2 ANOVA was conducted with Experiment (1B, 2A) and Encoding (Aloud, Silent) as between-subjects factors. Unlike with Experiment 1A, the effect of Experiment was not significant, $F(1, 104) = 2.41$, $MSE = .02$, $p = .12$. This is likely due to participants in Experiment 1B performing better than participants in Experiment 1A, bringing their average scores closer to those seen in Experiment 2A where all participants read the article twice. There was a significant effect of Encoding, $F(1, 104) = 8.91$, $MSE = .02$, $p, < .01$, $\eta_p^2 = .08$, $1\text{-}\beta = .84$, but again the interaction of Experiment and Encoding was not significant, $F(1, 104) = 2.99$, $MSE = .02$, $p = .09$.

Pure list designs often result in a smaller production effect, but Experiment 1B also failed to find a production effect, so it is likely that something about reading the article silently first contributed to the production effect observed here in Experiment 2A. Possibly, reading the article silently first and then out loud might have resulted in a sort of mixed list situation, albeit blocked, that highlighted the aloud-silent contrast. This in turn may have resulted in the use of a distinctiveness heuristic, in which participants remembered having read the materials out loud despite having also read them silently. Relatedly, when the materials are in the form of an extended text, production might be more effective for materials with which one is already familiar. It is also possible that reading the text first silently and then aloud resulted in strengthening of the text more than having read it twice silently, as the change in tasks may have drawn more attention to the unusualness of the task and hence to the importance of the text (McDaniel

& Bugg, 2008). Importantly for Experiments 3 through 5, the results of Experiment 2A indicate that a production effect can be obtained when texts have previously been read silently.

## Experiment 2B

Experiment 2B was conducted to investigate an issue not considered in Experiment 2A. In Experiment 2A, all participants read the article silently first; then, on a second reading, they read it either out loud or silently. However, there was no group that read the article out loud first and then read it silently the second time. It is possible that the presentation order of Aloud and Silent readings may be of importance. On the one hand, as long as the text is read in both modalities, there may be a benefit from producing on one of the two occasions, regardless of order. On the other hand, if the Aloud reading is second, this may result in greater distinctiveness than if the Aloud reading is first, as the prior Silent reading would provide a source of contrast, hence making the text more distinctive. Experiment 2B investigated whether an initial Aloud reading followed by a Silent reading would also result in a production effect.

### Method

**Participants.** Sixty-two undergraduate students from the University of Waterloo (41 female, mean age = 19.34, *SD* = 2.02) participated for course credit. From the Aloud group, data from three participants were excluded due to them reading silently when they were supposed to read out loud; four participants were excluded for skipping portions of the text; one participant was excluded due to reading too fast on the silent second reading (cutoff of 4.5 minutes determined by histogram analysis); data from two participants were excluded due to them writing during the reading phase (determined by audio inspection). From the Silent group, data from one participant were excluded due to reading audibly; two participants were excluded due to reading too fast (cutoff of 4.5 minutes determined by histogram analysis); and data from one participant were excluded due to accidentally skipping page 3 of

the text (pressed the "next page" button accidentally). This left 48 participants, 24 in each of the Aloud and Silent groups.

**Materials.** The materials were identical to those used in the preceding experiments.

**Procedure.** The procedure was identical to that of Experiment 2A, except that the Aloud group read the article out loud initially, and then read the article silently on the second reading whereas for the Silent group, both readings were done silently. As in the previous experiments, both stringent and lenient scoring procedures were conducted. Again, the pattern of data did not differ (with the lenient scoring resulting only in overall greater test performance), so the results are reported for the stringent scoring only.

**Results and Discussion**

*Reading Times*

In terms of reading times during the study phase, on the first reading, as expected, there was a significant difference between the groups: The Aloud group ($M$ = 765.73 s, $SD$ = 86.19) took longer to read the article than did the Silent group ($M$ = 644.98 s, $SD$ = 225.21), $t(29.60) = 2.45$, $p < .05$, equal variances not assumed. These data closely resemble those of Experiments 1A and 1B. Surprisingly, on the second reading, done silently for both groups, the Aloud group ($M$ = 484.02 s, $SD$ = 136.24) read the article faster than the Silent group ($M$ = 624.39 s, $SD$ = 247.40), $t(35.78) = -2.44$, $p < .05$. This may indicate that participants who read the article out loud initially may have felt more confident in their comprehension of the material (see Castel, Rhodes, & Friedman, 2013), and thus may have read through the article faster than they otherwise would have on the second reading.

*Memory Test: Fill-in-the-blanks*

The memory performance data for Experiment 2B are presented in the fourth row of Table 2. An independent samples *t*-test was conducted for memory performance on the fill-in-the-blanks test. The groups did not differ significantly, $t(46) = .51$, $p = .61$, $d = .15$, $1\text{-}\beta = .08$. From the Aloud group, 13 participants performed numerically greater than the Silent average, compared to 11 from the Silent group who performed greater than the Silent average.

Strangely, inspecting the data in Table 2, it appears that the largest difference in data between Experiment 2A and Experiment 2B was in the Silent conditions; the performance of the Aloud groups was comparable across the experiments. This is especially odd because the only difference between the experiments was a change that did not affect the materials, instructions, or procedure in the Silent group: They still read the text silently both times. This may be due to sample differences, as Experiments 2A and 2B were conducted at different times of year. Regardless, there would be value in future replications of production in rereading.

*Combined Analysis: Experiments 2A and 2B*

These findings suggest that the order of reading is important, as the production effect obtained in Experiment 2A was not obtained in Experiment 2B, despite the only change being the reversed order of aloud and silent reading. This is supported by a 2 x 2 ANOVA with Experiment (2A, 2B) and Production (Aloud, Silent) as between-subjects factors. Although the effect of Experiment was not significant, $F(1, 92) = 1.00$, $MSE = .02$, $p = .32$, there was a significant effect of Production, $F(1, 92) = 6.83$, $MSE = .02$, $p < .05$, $\eta_p^2 = .07$, $1\text{-}\beta = .74$, and a marginally significant interaction of Production and Experiment, $F(1, 92) = 3.58$, $MSE = .02$, $p = .06$, $\eta_p^2 = .04$, $1\text{-}\beta = .47$. Although the interaction did not reach significance, this may be because the Aloud scores in both experiments were numerically higher than the Silent scores.

It may be the case that the aloud group, having spent less time reading the text during the second, silent reading, consequently paid less attention to the text the second time. This pattern was

not noted in Experiment 2A and could indicate that reading text out loud may cause participants to become confident in their memory for the text and affect their motivation to read the text again silently. Importantly, though, having now shown that a production effect can be obtained when an article is read twice, with the second reading aloud, the critical experiment investigating the effects of choice while producing was conducted next.

## Chapter 3: Producing and Choosing

### Experiment 3

Experiment 3 addressed the central question of this dissertation: Does choosing what to read aloud influence memory? As laid out in the introduction, compared to having no choice, having the freedom to choose what to learn has been shown to have a beneficial effect on memory (Perlmuter et al., 1971; Watanabe, 2001) under other circumstances. Research on the production effect has thus far exclusively examined situations where participants have had no choice as to what they are to encode: They are given a list of words and told which words to read aloud and which words to read silently, with how to respond signaled for each individual word. However, in a realistic setting, people not only study different kinds of material, but often choose what they want to study and what they want to emphasize while studying. If we are to extend the production effect to the realm of studying, we must also investigate the effect of choosing what to produce while studying.

To investigate choice, participants in Experiment 3 were assigned to one of three groups: the Choice group, the Force group, or the Control group. All participants read the text once silently (as in Experiment 2A) and then were randomly assigned to group. During the second reading, participants in the Choice group chose which portions of the text to read aloud. Participants in the Force group were yoked to the Choice group, such that on their second reading each of them read the same text aloud as did a participant in the Choice group. To provide a baseline, participants in the Control group read everything silently again on the second reading. Would actively choosing what to read aloud lead to better memory than reading aloud sections that someone else had chosen? Although academic texts are largely unexplored in production effect literature, it was hypothesized that both the Choice and Force groups would show production effects, and that the ability to choose would lead to improved performance and less cost.

## Method

**Participants.** 141 participants from the University of Waterloo took part for course credit (98 female, mean age = 20.19, *SD* = 2.28). All were native English speakers and participated for credit.

It was decided to ask participants in the Choice group to choose between 25% and 33% of the text to read aloud. This decision was based on not wanting participants to be reading too much of the text aloud, requiring them to make real choices. Having read the text silently the first time, they could make informed choices on their second reading.

One concern with allowing participants to choose was that, whenever participants are given more freedom in how to accomplish a task, experimental control is weakened. Additionally, because participants in the Force group were yoked to those in the Choice group, any exclusions of participants in the Choice group required corresponding exclusions from the Force group. As a result, a considerable number of participant exclusions and replacements were expected. In the end, 51 participants were excluded. From the Choice group, data from one participant were excluded because the audio did not record; one participant was excluded due to closing Audacity; one participant was excluded for failure to retrieve the experimenter after the study phase, staying in the experiment room for a considerable length of time before taking the test; one participant was excluded for reading out loud in the silent reading phase; one participant was excluded for reading aloud paraphrases of the text instead of the actual text; one participant was excluded for skipping portions of the text while reading out loud; 12 participants were excluded for reading too little of the text out loud (less than 25%; an 8% buffer as participants could not be reasonably expected to know when they have read 33% of the text); and two participants were excluded for writing during the reading phase (determined by audio analysis). From the Silent control group, data from one participant were excluded because audio failed to record; one participant was excluded due to closing Audacity; one participant was excluded for failure to retrieve the

experimenter; two participants were excluded due to having left the experiment room; two participants were excluded due to reading out loud; one participant was excluded as the participant reported having read the article an additional time; and one participant was excluded due to writing during the reading phase (determined by audio analysis). From the yoked Force group, data from two participants were excluded due to experimenter error in underlining the portions of the text to be read out loud; three participants were excluded due to quietly mumbling the Aloud portions of the text; two participants were excluded for reading out loud some portions of the text that were to be read silently; one participant was excluded due to skipping portions of the Aloud text; two participants were excluded for rereading portions of the Aloud text; and one participant was excluded for writing during the reading phase (determined by audio analysis). Finally, in the Force group, an additional 10 participants were excluded due to being yoked to participants in the Choice group whose data were excluded; and one participant's data were excluded due to being an extra participant accidentally run through the experiment. This left a total of 90 participants: 30 in each of the Choice, Force, and Control groups.

**Materials**. The materials used in Experiments 1 and 2 were also used for Experiment 3, with the exception that the text was presented as paper printouts compiled in a binder, rather than via a computer screen. This permitted marking the portions to be read aloud in the Force group.

**Procedure**. The procedure was largely identical to Experiment 2A, except for the following modifications. First, the text was presented on four pages of paper instead of through E-Prime, and participants in all groups read the text silently the first time. On their second reading, participants in the Choice group were told to choose between one third and one half of the text material to read out loud. On their second reading, each participant in the Force group was given a copy of the article in which the parts read out loud by their yoked participant in the Choice group were already underlined. This was

accomplished by listening to the recordings of the Choice group. Participants in the control group read the text silently both times.

**Results and Discussion**

To determine performance as a function of Aloud vs. Silent within-subject, the first task was to identify which test questions corresponded to sections of the article that were read aloud vs. silently for each of the Choice participants. The sorting of the questions as Aloud vs. Silent was necessarily the same for the Force group due to yoking. To make the comparisons consistent, this sorting was also treated as the same for the Silent control group. Thus, whatever the Choice group read out loud (or silently) was treated identically for all three groups for the purpose of analysis. This resulted in participants being yoked into sets of three: one Choice participant, one Force participant, and one Silent participant per set.

The following analyses were performed on these 30 sets of participants. Scoring of the test answers was done in the same way as in Experiments 1 and 2. Because the stringent and lenient scoring patterns again did not differ (except that the lenient scoring produced overall greater performance), only the data from the stringent scoring procedure are reported. It is also worth noting that the proportion of fill-in-the-blanks test questions corresponding to text read aloud was .53 and the proportion corresponding to text read silently was .47. This indicates that, while reading, participants were not able to determine which parts of the text were going to be tested. It is also consistent with Choice participants being quite good at picking out the important, or to-be-tested, material to read aloud, since they were to read only about one third of the material, but this captured over one half of the test questions.

**Table 3.** Proportion of Correct Responses as a Function of Condition in Experiment 3

| | Aloud | Silent |
|---|---|---|
| Choice | .79 (.03) | .60 (.03) |
| Force | .71 (.04) | .55 (.04) |
| Control | .66 (.04) | .66 (.04) |

Note: Standard errors are in parentheses.

It is important to note that the all-silent Control group also had "aloud" data due to yoking, despite having never read the article out loud. Given that the Aloud/Silent designation was artificial in this group, there should have been no difference there; as Table 3 shows, performance of the Control group was identical in these two conditions.

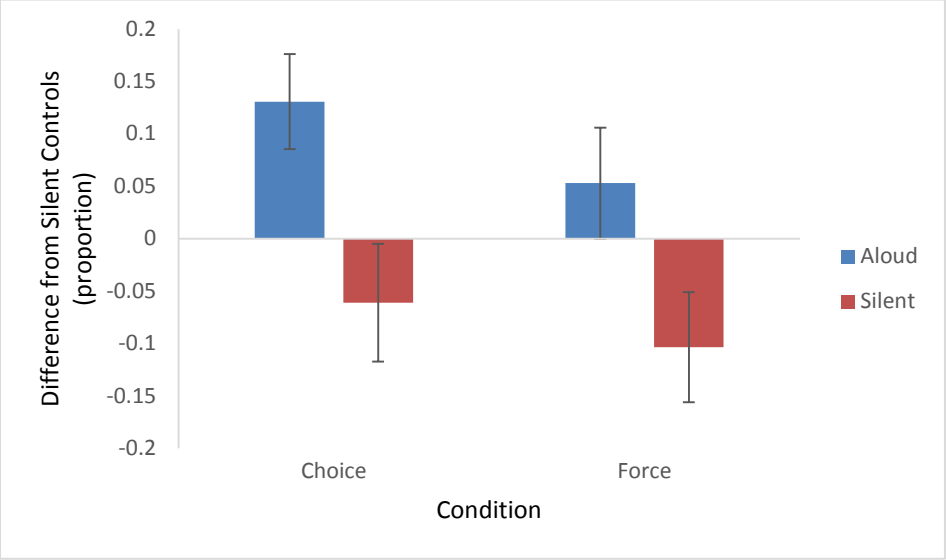
*Memory Test: Fill-in-the-blanks*

A 3 x 2 mixed ANOVA was conducted with Group (Choice, Force, Control) as the between-subjects factor and Encoding (Aloud, Silent) as the within-subject factor. There was a significant main effect of Encoding, $F(1, 87) = 38.40$, $MSE = .02$, $p < .001$, $\eta_p^2 = .31$, $1\text{-}\beta = 1.00$, reflecting a robust production effect: There were more correct answers to questions for which the text was read aloud ($M = .72$, $SD = .20$) than to questions for which the text was read silently ($M = .60$, $SD = .20$). The main effect of Group was not significant, $F(2, 81) = .87$, $MSE = .06$, $p = .42$, with The Choice group ($M = .69$, $SD = .15$) only slightly outperforming the Force ($M = .63$, $SD = .19$) and Control ($M = .66$, $SD = .19$) groups overall.

The key result was the significant interaction of Encoding with Group, $F(2, 87) = 9.94$, $MSE = .02$, $p < .001$, $\eta_p^2 = .19$, $1\text{-}\beta = .98$. Compared to the Silent control group, the Choice group showed a significant .13 benefit, $t(58) = 2.65$, $p < .05$, $d = .68$, $1\text{-}\beta = .74$, and a non-significant .06 cost, $t(58) = -1.20$, $p = .24$ , whereas the Force group showed a non-significant .05 benefit, $t(58) = 1.00$, $p = .32$, and a marginally significant .10 cost, $t(58) = -1.88$, $p = .07$, $d = .49$, $1\text{-}\beta = .46$. Thus, as can be seen in Figure 1

(which shows the proportion of difference from the Silent controls), the production effect was driven largely by a benefit to Aloud in the Choice group but by a cost to Silent in the Force group. From the Choice group, 23 participants showed a positive change from Silent controls on Aloud items (benefit) and 17 showed a negative change from Silent controls on Silent items (cost). From the Force group, 22 showed a benefit to Aloud items, and 19 showed a cost to Silent items. However, as Figure 1 shows, the difference was in the magnitude of costs and benefits, not in the number of participants showing them. From the Silent controls, 15 showed benefits to "Aloud" items and 15 showed costs to Silent items.

**Figure 1.** Baseline Corrected Costs and Benefits in Experiment 3




A production effect was obtained in both the Choice and Force groups, consistent with the findings of Ozubko et al. (2012). The equivalent performance of "Aloud" and "Silent" items in the silent Control group indicates that Choice participants are not simply selecting text on the basis of what is easier to remember. As Figure 1 shows, compared to the Silent group, the production effect in the Choice group was driven more by benefit than by cost whereas the production effect in the Force group was driven more by cost than by benefit. This is consistent with a lazy reading account, in which participants in the Force group are paying less attention to anything that is not underlined (indeed,

some participants whose data were excluded had skipped non-underlined passages altogether). Participants in the Choice group would be unable to ignore silent text, as they had to choose what to produce and what not to produce. The absence of a cost to non-produced text in the Choice group would be less consistent with an item-order account of cost, unless deliberation among items insulates them from disruption of order information.

The design of Experiment 3 has one potential weakness: It is not possible to determine whether the benefit to the Aloud items in the Choice group is due to the act of choosing or to a combination of production and choice. It is unlikely to be due entirely to production given that the Force group represents production without choice but the Force group did not outperform the silent Control group. Experiment 4 will address this issue.

## Experiment 4

Experiment 3 showed that the production effect is found when participants are both given and not given the option to choose what to produce, and that choice significantly influenced the pattern of cost and benefit. As laid out in the introduction, however, choice has also previously been shown to affect memory. In light of that, it is difficult to ascertain whether the benefit to the Aloud items in the Choice condition is due to choosing or to producing. The observed mnemonic benefit could be due, partially or entirely, to the act of choosing. As for the Force group, it seems likely that the cost could be due to "lazy reading" or possibly to some other cost related to not producing, such as disruption of order information (Forrin & MacLeod, 2016).

Experiment 4 was conducted in an attempt to disentangle the effects of production and choice. Experiment 3 had used only three conditions: Choice-Aloud (the Choice group), Force-Aloud (the Force group), and Force-Silent (the Control group). What is missing, therefore, is a Choice-Silent condition—a

condition in which a choice is made but no vocal response (including whispering and mumbling) is given, determined by audio analysis.

Unfortunately, using the same text materials as in Experiments 1 through 3 would not work well in creating a Choice-Silent condition: With those materials, it would not be feasible to include a condition where one could choose without producing their choice. The approximately 2000-word article is long, and keeping participants attentive with a long article can be a potential issue. Moreover, without vocal responses, it would be difficult to yoke participants in a Choice-Silent condition to those in a Force-Silent condition. Likewise, having participants underline the text that they choose may have unintended consequences, as underlining likely has its own effect (Hartley, Bartlett, & Branthwaite, 1980). A simple overt, but non-vocal response would be the best option, and a button press might be the best method currently available to achieve this. As it would be difficult if not impossible to map participants' arbitrary choices within full text articles using simple buttons—especially given that participants in Experiment 3 did not always choose entire sentences to produce—it was decided that switching to a different set of materials would be necessary. To address these issues, in Experiment 4, sentences replaced the text article, and in all cases Choice decisions were made through key press.

It was predicted that production and choice might interact such that the use of both techniques might lead to better remembering. Furthermore, it was predicted that being able to choose would be associated with less cost to silent items. Furthermore, it was expected that response times and false alarms would reflect this benefit, such that production and choice would be associated with faster responding and fewer false alarms on recognition tests.

**Method**

**Participants.** 113 participants from the University of Waterloo took part for course credit (91 female, mean age = 19.76, *SD* = 3.42). From the Choice-Aloud group, data from two participants were

excluded due to them not having produced the chosen sentences. From the Choice-Silent group, data from two participants were excluded due to audibly reading the text. From the Force-Aloud group, data from four participants were excluded due to failure to read the text audibly (determined by audio analysis); one participant was excluded due to reading out loud during initial presentation of sentences; one participant was excluded for singing "Somewhere Over the Rainbow" instead of reading sentences out loud; and two participants were excluded due to their corresponding participants in the Choice-Aloud group being excluded. From the Force-Silent group, data from three participants were excluded due to reading audibly, and two participants were excluded due to their corresponding participants in the Choice-Silent group being excluded. This resulted in 96 participants, 24 in each of the four conditions. Each participant in the Force-Aloud and Force-Silent groups was yoked to a participant in the corresponding Choice-Aloud and Choice-Silent groups.

**Materials**. The sentences used were those used in Ozubko et al. (2012). From their list of 548 sentences, 240 sentences were randomly selected. Of these, 160 were used to construct 80 sentence pairs (for the Choice procedure) with the remaining 80 sentences reserved to serve as lures on the recognition test. To control for item order across yoked Choice-Force participant pairs, the same materials were presented in the same order for the first 32 participants, then the list and order were randomized again for the next 32 participants, and again for the last 32 participants. The experiment was conducted in PsychoPy3 (Peirce et al., 2019).

**Procedure**. Participants sat at a computer and were told that they would be presented with a series of pairs of sentences, with the two members of a pair presented successively. They were told that there would be a memory test at the end of the study phase. Each sentence appeared on the screen for 3.75 s, with either "[1]" or "[2]" appearing to the left of it. After both sentences had been presented and removed from the screen, participants were given instructions depending on the condition to which

they were assigned. Participants in the Choice-Aloud group were told to press "1" or "2" to choose the sentence that they thought might be more likely to be tested. This sentence would then be re-presented for them to read aloud. Participants in the Choice-Silent group were given the same instruction, but told to read the chosen sentence silently when it appeared. Participants in the Force-Aloud group, after reading both sentences, were shown one of the sentences—the sentence chosen by the yoked Choice participant—and told to read it aloud. Participants in the Force-Silent group were given the same instruction, but told to read the re-presented sentence silently.

Participants pressed 1 or 2 to make their selection (Choice condition) or pressed 1 or 2 to move on to the preselected sentence (Force condition). The chosen/forced sentence was then presented for 3.75 s and the participant read it aloud or silently as appropriate for their condition. This was repeated for each of the 80 sentence pairs, for a total of 160 sentences. In the recognition test that followed study, all 160 old sentences were presented, intermingled randomly with 80 new sentences. Participants were told to press "M" if the sentence was old or to press "C" if the sentence was new.

**Results and Discussion**

**Table 4.** Hits and False Alarms in Experiments 4 and 5

| | | Aloud | | | Silent | | |
|---|---|---|---|---|---|---|---|
| | | Hits | | False Alarms | Hits | | False Alarms |
| | | Chosen | Non-chosen | | Chosen | Non-chosen | |
| Experiment 4 | Choice | .83 (.02) | .61 (.03) | .07 (.01) | .80 (.02) | .68 (.02) | .09 (.02) |
| | Force | .78 (.03) | .51 (.03) | .12 (.02) | .67 (.02) | .55 (.03) | .14 (.02) |
| Experiment 5 | Choice | .83 (.02) | .63 (.03) | .11 (.02) | .75 (.03) | .61 (.03) | .11 (.02) |
| | Force | .79 (.02) | 58 (.02) | .16 (.02) | .74 (.02) | .60 (.03) | .16 (.02) |

Note: Standard errors are in parentheses. In Experiment 5, Aloud and Silent false alarms are identical.

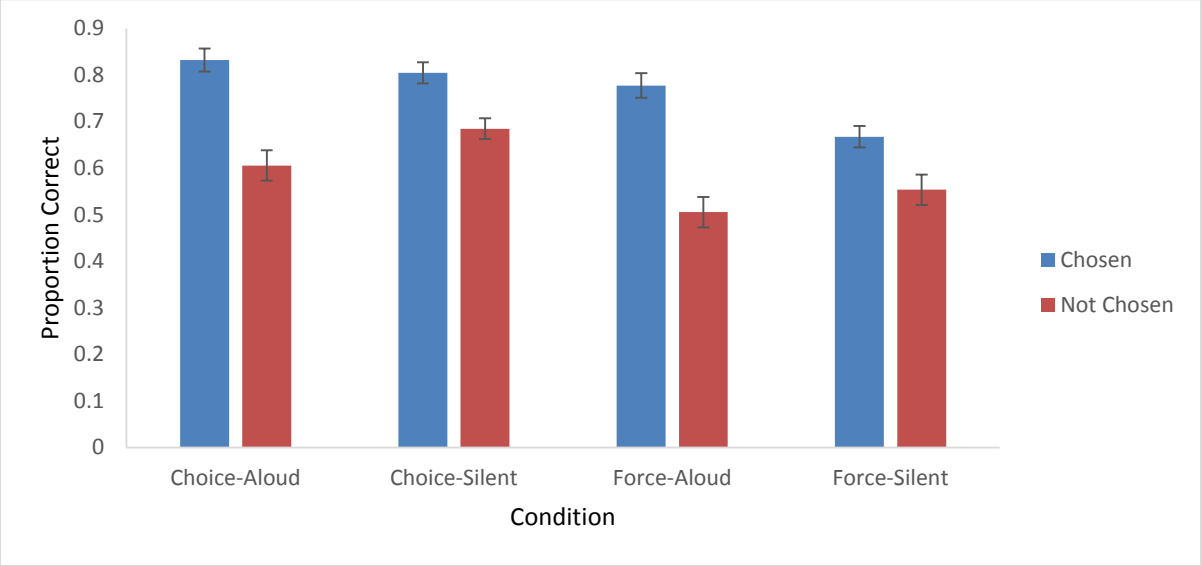
*Memory Test: Hits*

A 4 x 2 mixed design ANOVA on recognition hit rate was conducted, with Condition (Choice-Aloud, Choice-Silent, Force-Aloud, Force-Silent) as the between-subjects factor and Selection (Chosen, Non-chosen) as the within-subject factor. Of note is that Selection, in the Aloud groups, acts as mixed list production, given that participants would read out loud the chosen sentence. There was a significant effect of Selection, $F(1,92) = 286.19$, $MSE = .01$, $p < .001$, $\eta_p^2 = .76$, $1\text{-}\beta = 1.00$, a significant effect of Condition, $F(3, 92) = 6.25$, $MSE = .03$, $p < .01$, $\eta_p^2 = .17$, $1\text{-}\beta = .96$, and a significant interaction, $F(3, 92) = 13.20$, $MSE = .01$, $p < .001$, $\eta_p^2 = .30$, $1\text{-}\beta = 1.00$, see Table 4 for hits and false alarms.

The Selection effect indicates that memory for the chosen sentences was better than memory for the non-chosen sentences. However, it is important to realize that, for all groups, the chosen sentences were presented twice whereas the non-chosen sentences were presented only once. The interaction between Selection and Condition is more interesting, reflecting at least in part the presence of a mixed list production effect when inspecting the difference between chosen and non-chosen sentences; the difference in recognition performance between chosen and non-chosen sentences was greater when the chosen sentence was produced, that is, in the Choice-Aloud and Force-Aloud groups (see Figure 2).

Post-hoc Tukey tests showed that the Choice-Aloud ($M = .72$, $SD = .13$) group performed better than the Force-Silent group ($M = .61$, $SD = .13$), $p < .05$, but not the Force-Aloud ($M = .64$, $SD = .13$) group, $p = .14$; interestingly, the Choice-Silent group ($M = .74$, $SD = .10$) performed as well as the Choice-Aloud group, $p = .89$, outperforming both the Force-Aloud, $p < .05$, and the Force-Silent, $p < .05$, groups. No other differences were significant.

**Figure 2.** Proportion of Hits as a Function of Condition in Experiment 4. Error bars represent ±1 standard error.




That the performance of the Choice-Silent group was equivalent to that of the Choice-Aloud group was somewhat unexpected, and the Choice-Silent group seems to show greatest recognition of non-chosen sentences; compared to the Choice-Aloud group, the benefit in memory for non-chosen sentences in the Choice-Silent group bordered on significant, $t(40.60) = 2.00$, $p = .052$, $d = .58$, $1\text{-}\beta = .50$. Importantly, planned comparisons showed that the Choice-Aloud group showed greater memory for non-chosen sentences than did the Force-Aloud group, $t(46) = 2.17$, $p < .05$, $d = .63$, $1\text{-}\beta = .57$. The results of this comparison of the Choice-Aloud and Force-Aloud groups indicate that choosing can protect against the cost of producing. Again, due to the what is essentially a confounding of Selection with repetition (chosen sentences were read twice whereas unchosen sentences were read only once), this is difficult to interpret, but it seems possible that the act of choosing insulated memory of the non-produced items from the cost typically seen in mixed list production, consistent with the findings of Experiment 3. The Force-Aloud group did not perform better than the Force-Silent group, despite memory for chosen sentences being greater in the Force-Aloud group than in the Force-Silent group, $t(46) = 3.12$, $p < .01$, $d = .90$, $1\text{-}\beta = .86$; this is likely due to the fact that the comparison also involved

memory for both chosen and non-chosen sentences, in which memory for non-chosen sentences was numerically lower in the Force-Aloud group than in the Force-Silent group.

From the Choice-Aloud group, 21 participants showed greater recognition for chosen sentences than the Force-Silent average, as did 21 participants from the Choice-Silent group. From the Force-Aloud group, 17 participants performed better than the Force-Silent average. From the Force-Silent group, 12 participants showed better performance than the Force-Silent average.

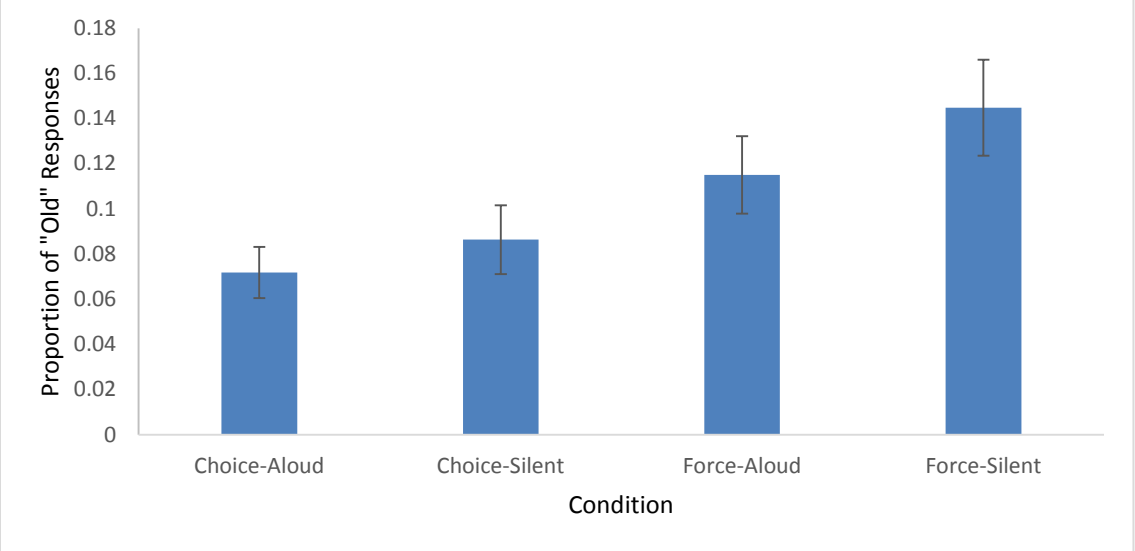
Now considering memory for sentences that were chosen or pre-selected, the Choice-Aloud, Choice-Silent, and Force-Aloud groups did not significantly differ (see Figure 2). Using the data from the Choice-Aloud and Choice-Silent groups, it was revealed that the current sample size yielded a power value of .13 to find a benefit of Choice-Aloud over Choice-Silent in chosen sentences. To reach a power value of .80 would require rather large samples of 280 participants per group. Both production and choice seem to have beneficial effects on memory for sentences, but the results of the current experiment do not suggest that there is an additional benefit to chosen sentences beyond production when both encoding manipulations are involved.

*Memory Test: False Alarms*

Next, the false alarms were investigated. Although choosing and producing did not interact to produce a significant benefit, it is possible that evidence of this benefit might be obtained in reduced false alarms. Produced items are more distinctive, and presumably items that are chosen have the distinctive quality of being chosen, and there is reason to believe that distinctiveness influences false alarms (Forrin et al., 2016). A one-way ANOVA conducted to examine the effect of Condition on false alarms was significant, $F(3, 92) = 3.77$, $MSE = .01$, $p < .05$, $\eta_p^2 = .11$, $1\text{-}\beta = .80$. As the results of Levene's test were significant ($p = .02$), Games-Howell post-hoc tests were used to account for unequal variances, with the Choice-Aloud group ($M = .07$, $SD = .06$) showing lower false alarms than the Force-Silent group

(*M* = .14, *SD* = .10), *p* < .05. However, although other comparisons were not significant, Figure 3 suggests the possibility of a trend where false alarms decrease as encoding strategies increase. This was investigated and a monotonic trend was found in false alarms as a function of Condition, *t*(42.34) = -3.27, *p* < .01, not assuming equal variances. This indicates that both choice and production reduce false alarms, supporting the hypothesis of both techniques imparting some level of distinctiveness to processed items. It is entirely possible that when items presented at test are new, participants may discriminate these items as neither having been produced nor deliberated on for prior choosing during study.

**Figure 3.** False Alarms as a Function of Condition in Experiment 4. Error bars represent ±1 standard error.




*Response Times*

Response times on the recognition test were also examined, and an effect of Selection was found, *F*(1, 92) = 17.79, *MSE* = 62,712.97, *p* < .001, $\eta_p^2$ = .16, 1-β = .99, indicating that chosen sentences (M = 1784 ms, *SD* = 459.59) were recognized faster than non-chosen sentences, (*M* = 1937 ms, *SD* = 553.37). This is

to be expected, of course, because chosen sentences received an additional rehearsal. No other response time effects were significant.

The results of Experiment 4 provide valuable insight into how both production and choice influence memory. As expected, both production and choice improved memory performance, yet any evidence that they can interact to produce an additional benefit was minimal. One issue with this experiment was that the Selection factor—whether an item was or was not chosen—was confounded with repetition: Chosen sentences were presented twice whereas unchosen sentences were presented only once. Silently read sentences would thus also have only been read once, whereas aloud sentences would have been read twice. As a result, interpretations of the mixed list production effect in Experiment 4 must be tentative. To address this, Experiment 5 replicated the procedure of Experiment 4, except that production was moved from a between-subjects variable to a within-subject variable; in this way, production could be manipulated separately from choice to prevent the repetition confound.

## Experiment 5

For Experiment 5, the procedure of Experiment 4 was used, with the key exception that production now was a within-subject factor. In Experiment 4, silent sentences were read only once instead of twice, making it difficult to interpret the cost of production. With production manipulated within-subject, it became possible to have data for silently read sentences in a mixed list that are independent of the repetition confound.

### Method

**Participants.** A total of 99 participants from the University of Waterloo (84 female, mean age = 20.19, *SD* = 3.53) participated for course credit. The target sample size was increased from 24 to 36, as half of the sentences for all participants were blocked under the Aloud instruction and the other half

were blocked under the Silent instruction, reducing the number of observations per condition. From the Choice group, data from two participants were excluded due to failure to read sentences out loud; three participants were excluded due to rereading sentences out loud; three participants were excluded due to reading audibly during the silent trials; one participant was excluded for reading out loud during initial presentation of the sentences; and one participant was excluded for presumably reversing the keys for "old" and "new", as indicated by very low hits (Aloud = .10, Silent = .30) and very high false alarms (.89). From the Force group, data from five participants were excluded due to failure to read sentences out loud; five participants were excluded due to reading out loud during initial presentation of the sentences; one participant was excluded due to reporting foreknowledge of the production effect; and four participants were excluded due to their corresponding participants in the Choice condition having been excluded. This left 35 participants in the Choice group and 35 in the Force group. Due to the COVID-19 pandemic resulting in the closing of research laboratories, it was not possible to recruit the last two participants.

**Materials.** The materials in Experiment 5 were identical to those of Experiment 4.

**Procedure.** Experiment 5 was identical to Experiment 4 in every respect except that production was changed to a within-subject variable; choice remained a between-subjects variable. After initial presentation of both sentences, participants had to choose a sentence to remember in Choice trials, or else a sentence was pre-selected for them on Force trials, and they then read the chosen sentence out loud or silently. To avoid overburdening participants with having to remember multiple instructions, production was blocked such that, for half of the participants, the first 40 sentence pairs were presented under Aloud instructions and the last 40 pairs were presented under Silent instructions, and the rest of the participants were presented with the Silent block first and the Aloud block second.

**Results and Discussion**

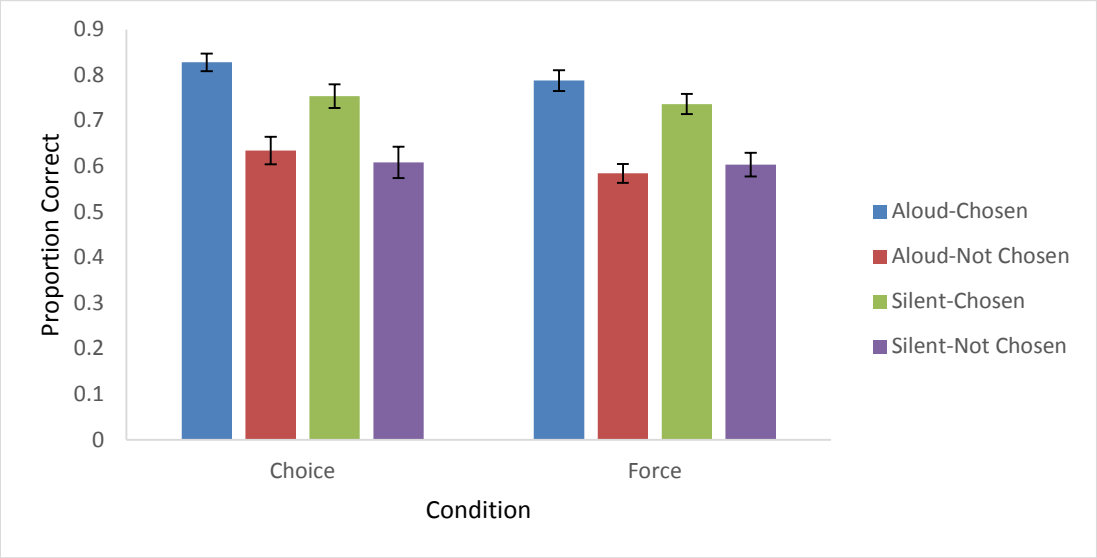
For the following analyses, the Selection variable from Experiment 4 indicating chosen or non-chosen sentences was removed: This was done for two reasons. First, with production moved within-subject, analysis of a mixed list production effect without the problem of repetition can be done just by examining chosen or pre-selected sentences, as participants read chosen sentences in both Aloud and Silent blocks twice. Second, including Selection in the analyses would mostly add noise, as non-chosen sentences from the Aloud blocks were never read out loud yet would still be considered as Aloud. Regardless, data for the non-chosen items can be found in Figure 4 and Table 4. Additionally, block order was included as a between-subjects factor, given that the results of Experiment 2A and 2B indicated that order of production when blocked may be of some consequence.

*Memory Test: Hits*

A mixed design 2 x 2 x 2 ANOVA was conducted on recognition hits with Condition (Choice, Force) and Order (Silent First, Aloud First) as between-subjects factors and Production (Aloud, Silent) as the within-subject factor. Figure 4 shows the descriptive statistics. The effect of Condition was not significant, with the Choice group ($M$ = .79, $SD$ = .12) performing no better than the Force group ($M$ = .76, $SD$ = .12), $F(1, 66) = .92$, $MSE = .03$, $p = .34$. There was, as usual, a significant effect of Production, $F(1, 66) = 19.00$, $MSE = .01$, $p < .001$, $\eta_p^2 = .22$, $1\text{-}\beta = .99$, indicating that Aloud items ($M$ = .81; $SD$ = .13) were better remembered than Silent items ($M$ = .75; $SD$ = .14). The only interaction that approached significance was that of Production and Order, $F(1, 66) = 3.35$, $MSE = .01$, $p = .07$, consistent with the possibility that production is more effective when the aloud block follows the silent block; this could of course simply reflect recency. None of the other interactions were significant, all $F$s < .70. From the Choice group, 26 participants showed Aloud recognition performance greater than the average of the Force-Silent items, and 22 from the Choice group showed the same pattern for Silent items. From the

Force group, 24 showed greater performance on Aloud items than the Force-Silent average, and 18 from the Force group showed the same for Silent items.

**Figure 4.** Proportion of Hits as a Function of Condition in Experiment 5. Error bars represent ±1 standard error.




*Memory Test: False Alarms*

A 4-level, one-way ANOVA investigating the effects of Condition on false alarms was non-significant, $F(1, 68) = 2.71$, $MSE = .01$, $p = .10$. In other words, participants in the Choice group did not perform any better in regard to false alarms than did participants in the Force group, contrary to the results of Experiment 4. This is consistent with the lack of effect of Condition on hits noted in the above paragraph. The ability to choose, in Experiment 5, apparently contributed no benefit to participant performance.

*Response Times*

In terms of response times, a 2 x 2 ANOVA with Condition (Choice, Force) as the between-subjects factor and Production (Aloud, Silent) as the within-subject factor found a significant effect of Production, $F(1, 68) = 20.07$, $MSE = .03$, $p < .001$, $\eta_p^2 = .23$, $1-\beta = .99$, indicating that sentences that had

been read out loud (*M* = 1615 ms, *SD* = 349.09) were faster to respond to than were sentences that had been read silently (*M* = 1736 ms, *SD* = 438.33). This is consistent with the benefit conferred on recognition by production. No other response time effects were significant.

Overall these results are surprising. It was anticipated that Experiment 5 would show similar results to Experiment 4, namely, benefits of both production and choice. The results of Experiments 3 and 4 clearly showed effects of choosing on memory, as does an increasing wealth of literature, yet Experiment 5 failed to find any effects at all of choice. Neither was there a benefit in false alarms for the Choice group. The only difference between Experiments 4 and 5 is that the instruction to produce was changed from a between-subjects variable to a within-subject factor. It is possible that this change negatively influenced the previously observed beneficial effect of choosing. This will be discussed at length in the General Discussion to which I now turn.

## Chapter 4: General Discussion

The series of experiments in this dissertation addressed how choosing what to produce influences memory. By now, production is well established as causing a memory benefit: Reading isolated items aloud improves memory for those items relative to items read silently. This is true when all of the items are read aloud as opposed to silently—the between-subjects, pure list procedure—and the effect is notably larger when only some of the items are read aloud—the within-subject, mixed list procedure. A growing body of evidence delineates the extent and the boundaries of the production effect (see MacLeod & Bodner, 2017, for a brief review). It is now time to push beyond simple isolated items and to examine the influence of production on more elaborate materials such as text, investigating the implications in "real world" situations.

Ozubko et al. (2012) demonstrated a production advantage for portions of text that were read aloud in a mixed list situation, but did not examine the pure list situation. Kline (2019) did investigate the pure list situation but found no evidence of a production effect, although she claimed that her study suffered from small sample sizes. Consequently, Experiments 1 and 2 here investigated pure list production with longer texts. Both studies from Experiment 1 established that, with longer narrative texts, a pure list production effect was absent. Experiment 2 qualified this result, first showing that a between-subjects production effect can be found when the aloud reading follows an initial silent reading (Experiment 2A), and then showing that this benefit vanishes when the aloud reading comes first instead of second (Experiment 2B). Experiment 2 also provided a bridge to Experiment 3 by having all participants do an initial silent reading before reading anything out loud.

The other central issue of this dissertation was the question of choice. In a real world setting, people typically read out loud content that they choose to read out loud. Experiments 3, 4, and 5 addressed this issue. Experiment 3 replicated the mixed list production effect of Ozubko et al. (2012),

while also showing that it can be obtained even when participants had previously read the materials silently (i.e., on second reading). That is, familiarity with the materials does not negate the production effect. In terms of the effects of choosing, both the Choice and Force groups yielded a production effect. Interestingly, compared to a silent control, the Choice group production effect was due more to a benefit to the aloud portions, and the Force group production effect was due more to a cost to the silent portions. Additionally, it seems that the silent text portions in the Choice group were protected from incurring a cost from production. This is understandable, given that Choice participants were required to deliberate over all the text to choose which parts to read aloud whereas the Force participants faced no such requirement.

Although Experiment 3 showed that production while studying is indeed a viable study strategy, this experiment lacked a choice-only condition. Experiment 4 addressed this issue with a simpler design, using sentence pairs and recognition testing. While both choice and production have been established to improve memory, choice may be associated with less cost when combined with production. Interestingly, combining choice and production did not increment the benefit beyond what either strategy would have imparted individually. False alarm data, however, suggested that both choice and production resulted in a reduced likelihood of mistakenly viewing new items as old. This is consistent with the use of a distinctiveness heuristic for both production and choice (see Dodson & Schacter, 2001), in which participants search memory for information regarding the status of new items, specifically whether they were previously read aloud or chosen. Production without choice in Experiment 4 showed a clear benefit to produced items, yet did not do so in Experiment 3. This difference may be due to the different test formats—fill-in-the-blanks vs. recognition—or possibly to the different types of materials, where one experiment used a long narrative text and the other a series of independent sentences. Experiment 5 presented a strange finding, namely, when production was switched to a within-subject

factor, the between-subjects effect of choice vanished—yet the effect of production remained. It is possible that one effect overshadowed the other, a possibility that will be discussed further.

## Theoretical Perspectives: Production and Choice

Of the two primary accounts of the production effect—distinctiveness and strength (MacLeod et al., 2010; Bodner & Taikh, 2012; for a review see MacLeod & Bodner, 2017)—the distinctiveness account appears to better fit the present data. In a strength account, the memory trace of a produced word is strengthened by production, whereas distinctiveness would suggest that produced words are better remembered because of the distinctive quality encoded with them, allowing greater discrimination. Experiment 1, which investigated pure list production, did not show any evidence of a production effect. If reading a text aloud was enough to strengthen memory of that text (relative to the same text read silently), one might have expected to see a production effect. Yet, even when combining the data from Experiments 1A and 1B, no production effect was found. In Experiment 2, where the text was read silently first and then aloud, a production effect was observed. If the contrast between having read it silently and then aloud could result in distinct aloud information being processed together with the text during encoding, this could result in a sort of mixed list situation, in which the production effect is ordinarily considerably more robust. Thus, the results of Experiments 1 and 2 more strongly support a distinctiveness account over a strength account.

Concerning choice, Experiment 3 has shown that, compared to a Control group, the Choice group was resistant to the cost of production. Experiment 4 also yielded evidence suggesting that the ability to choose not only can aid memory for non-chosen alternatives but also can protect from the cost of production to silent items. This is consistent with the literature showing that choice imparts benefits to non-chosen items (Monty et al., 1982; Watanabe, 2001; Watanabe & Soraci, 2004), and is also consistent with Driver's (2013) two-stage process account of choice, in which participants first

deliberate upon items and then select one option, where both deliberation and selection have their own mnemonic consequences. Although the design used in these experiments did not allow for the possibility to investigate deliberation and selection separately, this would be a valuable direction for future research.

## Overshadowing

As mentioned previously, in Experiment 5 the choice effect vanished. Experiment 5 was identical to Experiment 4 in all ways except that production was manipulated as a within-subject factor. Odd as it may be, it is not unheard of for one strategy to overshadow another. Putnam et al. (2014), in investigating the production effect in associative memory, had participants learn word pairs either by reading them silently, typing them, or saying them out loud. In their initial experiments, after each pair, participants underwent a semantic relatedness task on the pairs to ensure that associative processing had taken place. However, with the addition of the semantic relatedness task, the production effect was eliminated. When the same procedure was implemented but without the semantic relatedness task, there now was a reliable production effect, both for individual items and for the associative link between pairs. It would appear that something about the semantic relatedness task had overshadowed production.

Unlike the findings of Putnam et al. (2014), in the current experiment it was the effect of production that remained but the effects of another strategy (choice) disappeared. This is similar to the findings of Cho and Feldman (2013) who found that production overshadowed another retrieval strategy. Participants in their experiments heard words in either familiar American or unfamiliar Dutch accents, and some of the words were produced. In their Experiment 1A, participants produced in their normal speech; in Experiment 1B, participants imitated the accent of the initial speaker of the words. There were production effects in free recall as well as greater performance on words that were

presented in a Dutch accent. However, in recognition, only words that were presented in the silent condition (in which participants only heard the speaker but did not produce) benefited from Dutch accents; in the production group, the benefit of accent had disappeared. Their Experiment 2 showed similar findings. Although they found that accent aided memory, this effect was considerably muted when production also took place. This was interpreted as participants favouring the use of production information over accent information for retrieval whenever production information was available.

Cho and Feldman (2016) replicated this overshadowing effect in production in experiments that used English and Chinese accents instead of English and Dutch accents. They again found that the benefit of accent in recognition was only present for non-produced words; recognition of produced words was not aided by the unfamiliar accent. Their Experiment 2 featured a source monitoring task that revealed that accent information was not lost even when its effects on recognition of produced words was not present. This indicates that accent information was simply not needed—or not used—when production information was available. It is noteworthy in this regard that Ozubko et al. (2014) showed that even when silent items were strengthened by repetition, such that their recognition equaled that of aloud items, participants could still discriminate the aloud items from the silent items.

Similarly to my findings, Taikh and Bodner (2016) found that when production was manipulated as a between-subjects factor and imagery was manipulated as a within-subject factor—in which some words were visualized in capital letters or the words' referents were visualized—the pure list production effect was not obtained, whereas an effect of imagery was found. It is likely that participants in these conditions used a strategy of recollecting imagery information instead of production information to make their judgments. It seems likely that when two manipulations that improve memory occur in tandem, one may overwhelm the other despite each ordinarily being beneficial on its own. Put another

way, the two manipulations will not have additive influences (although see Forrin, Jonker, & MacLeod, 2014, for a counterexample).

It is possible that something similar occurred in my experiments. In Experiment 4's Choice-Aloud group, the chosen aloud sentences were presented more times than the non-chosen silent sentences and the chosen sentences were also always aloud sentences. Production and choice were not separated within the group. These factors may have emphasized the act of choosing. However, in Experiment 5, the production task was manipulated within-subject and made especially salient. Participants were required to switch tasks in the experiment, drawing attention to the production nature of the task, unlike in Experiment 4. Hence, on the subsequent recognition test, participants might have been motivated to ask themselves "do I remember saying this?" instead of "do I remember choosing this?" It is even possible that there is a kind of "pecking order" for manipulations, with those higher in the order overruling those lower in the order when they co-occur. It is possible—assuming that production overshadowed choice due to being more salient, or that the design may have encouraged participants to opt for one strategy over the other—that benefits of both production and choice might be preserved if the two techniques were used separately. In this way, the two manipulations would not be competing with each other. If participants chose first and after all choices are made they then engaged in production, it is possible that effects of both manipulations would be evident. This would, however, more likely be the case if production involves a significant strength component, as a strengthened memory trace need not rely on a distinctiveness heuristic to be successfully remembered.

## Real World Applications

What does this all mean for the studier? How can one enhance studying? The findings of Experiments 1 through 3 inform us as to how one may wish to study. For situations in which a study guide is available, the studier already knows what is important and would not need to worry about what

to choose. In these situations, one may decide to read all of the relevant course content in a given text and to read the content specified by the study guide out loud. Essentially, this would constitute a mixed list scenario, and the benefits of production for such a situation have already been established by Ozubko et al. (2012) as well as in the current experiments. Alternatively, as I have shown, a production effect may still be obtainable even if the same materials have been read twice, once silently and once aloud—a novel result that may have widespread applications. In the case of studying, one could read the to-be-tested content silently while studying, then read it again out loud. Both scenarios appear analogous to the mixed list situation seen in laboratory studies using production.

When a study guide (or other index of what is important) is not available, it will be up to the studier's own judgment to determine what content is important. Experiment 3's results would suggest that, when it comes to studying, making a conscious effort to choose what is important may matter more than vocal rehearsal. The choice effect is well established to aid memory, and memory for both chosen and non-chosen items is enhanced (Monty et al., 1982; Watanabe, 2001). Choice could make up for any cost incurred by production. A studier could then choose what content is important, and read that content aloud. Indeed, even without worrying about the "lazy reading" account (Begg & Snider, 1987), production might still result in a cost according to an item order account (Jonker et al., 2014; Jonker & MacLeod, 2015; McDaniel & Bugg, 2008). Yet there is little reason to think that the choice procedure would disrupt item order in a list as no item is processed in an unusual way, and choice seems very unlikely to result in lazy reading as the reader must consider all options. Of course, it is possible that just choosing is itself enough, and that production would not impart as large a benefit.

Although choice may seem to be a reliable method of studying, it is not without limitations. For the full benefit of choice, the chosen items must correspond with the tested items. In a situation where the items selected by the studier are not the items that are tested, the benefit of choice may be

minimal. Considering Driver's (2013) stages of deliberation and selection, there would still be a benefit of deliberation, but the benefit of selection would be absent. How much of the total benefit of choice would remain intact in such a situation would be difficult to ascertain with the data currently available. If it is the case that choice imparts minimal or no benefit when the studier fails to correctly determine what content will be tested, it may be better simply to rely entirely on production. In such a situation, where one does not know what will be tested and is not confident in their ability to select important materials from a text, it may be more efficient to vocally produce the text. Although pure list production does not seem to benefit studying for a single reading, reading everything silently first then reading everything aloud does seem to have a beneficial effect on memory during tests.

Of course, there is no reason that the above techniques cannot be combined in other ways. One could always produce the text while studying, then subsequently read through the text and choose which items may be important and vice versa. Both production and choice are very simple study strategies, both of which are beneficial and do not require any special preparations to engage in. It is possible that using both techniques may cause one to overshadow another but it is also possible that, as long as the strategies are not made to compete during retrieval, the studier may benefit fully from both production and choice.

### Limitations of the Current Experiments

The current experiments are not without limitations. Although Experiment 1 has shown that a pure list production effect was not present in studying longer texts, Experiment 2 qualified that result by showing that reading a text silently initially and then subsequently reading it aloud can result in a production effect. However, when the text was read aloud first and then silently second, there was no production effect. It is possible that having read the text aloud the first time, being told to read it silently a second time might be tedious, or participants felt less motivated to read with the same diligence on

the second occasion; this is supported by response time data seen in Experiment 2B. Participants in Experiment 2A—in which they read silently first and aloud second—could not afford to give the text less attention on the second reading, as they were required to read every word out loud. However, if someone is studying a text such that they read the same text both aloud and silently, order may not matter as much as these experiments make it seem. If a studier endorses such a strategy, they may be doing so out of a motivation and willingness to read the text twice. In such a situation, even if they read the text aloud first, they may be more inclined to devote as much attention during a second reading, even when the second reading is silent. In such a situation, it is possible that they would benefit from production. This possibility is yet untested.

Experiment 3 has shown that production with choice while studying is superior to production without choice. However, Experiment 3 was unable to test choice without production. The reason for this is that it would be difficult to yoke a Force group to a Choice group without any observable behaviour to yoke them by. Ideally, for the purposes of this dissertation, it would have been best to have a Choice-Silent group in Experiment 3, but attempting that would result in having no indication of what participants in that group chose. Experiment 4 partially addressed this concern with a simpler design that allowed participants to choose one of two sentences by keypress, in both Silent and Aloud conditions, but Experiments 3 and 4 differ in some other significant respects. Experiment 3 used a long, narrative text about toucans, and participants were tested on fill-in-the-blank questions. Experiment 4 used a series of independent sentence pairs, and participants were tested on recognition. The experiments differed largely in the materials used and the test format, and these may have also affected the pattern of results.

Whereas Experiment 3 showed that production without choice resulted in mostly cost, Experiment 4 showed a benefit as well. This may be a feature of test format, as in production effect

research recognition is known to show a production benefit (Hopkins & Edwards, 1972; Ozubko & MacLeod, 2010; Forrin et al., 2016) whereas free recall is known to show only costs in a mixed list design (Jones & Pyc, 2014; Lambert, Bodner, & Taikh, 2016; Forrin & MacLeod, 2016). Fill-in-the-blanks testing in this case seems to align more with the pattern of results seen in recall testing. As a result, the findings in Experiment 3 may only be directly generalizable to test formats that use fill-in-the-blanks tests. Other test formats, such as short essays and multiple choice questions, have not been investigated. Another issue is ensuring compliance in silent reading conditions. Clearly, it is difficult to ensure that participants are actually reading when there is no observable behavior. Eye-tracking may be viable in future studies, but for the purposes of this dissertation—which sought to emulate a realistic scenario—using eye-tracking would have been rather unlike standard reading.

## Future Directions

Naturally, a good extension of this line of research would be to examine the production and choice effects in a design similar to Experiment 3 but with a multiple choice test instead. If recognition tests can result in a benefit to produced items, as Experiment 4 has shown, then it is possible that, when choice is not involved, a multiple choice test might show benefits of production instead of just costs. Accordingly, investigations of production and choice can be done with any relevant test format, and doing so will help to reveal which strategy may be best for any specific test format.

Experiment 2 has shown that the same text can benefit from being read aloud after being read silently as opposed to being read silently twice. As of yet, no such effect has been found in the word-based production effect literature. It would be a worthwhile endeavor to investigate this possibility under typical production effect laboratory conditions, with lists of isolated words. If it is the case that a production effect is still found, this could very well change how researchers view the production effect and its applications in the real world. Memory of an item may thus be considered to be distinctive not

only in relation to other items lacking that distinctive quality, but even in relation to memory of the same item processed differently. Could an individual remember a date better if they silently rehearsed that date, then spoke it aloud, compared to silently rehearsing it twice? Could the same be true for remembering phone numbers or other complex combinations? Importantly, such a method of production would logically eliminate the associated cost to silent items, as records of both silent and aloud encodings would refer to the same item. This method may be superior to typical pure list production in regard to eliminating cost and retaining a benefit.

Concerning the possibility of overshadowing of one encoding dimension by another, this is an issue that could affect not only production and choice but many other encoding and retrieval strategies under certain conditions. As a result of these potentially widespread implications, the conditions under which overshadowing can occur is something that ought to be investigated. If it is actually the case that having two techniques compete against each other may encourage a participant to use only one of the techniques for retrieval, it is possible that using different strategies to encode the same items can be more effective if they are done independently of each other. This would especially be the case if participants decide to focus on one strategy during encoding and to neglect the other.

Although the current dissertation has investigated production and choice, and presupposed Driver's (2013) two-stage process account of choice as an explanation for mnemonic benefits seen for non-chosen items, the current experiments have not investigated the distinction between deliberation and selection. In fact, surprisingly, this distinction of the two possible sub-processes of choice has received very little attention. Researchers in the future may want to consider investigating what kinds of benefits are conferred by these processes, and how much of any given benefit is due to deliberation and how much is due to selection.

## Conclusion

In sum, the experiments here have built upon the findings of Ozubko et al. (2012) and have further shown that under some conditions—conditions that do occur in normal studying—production benefits learning in long texts. The effect is most prevalent in mixed list situations, when some of the material is studied aloud and some is studied silently, but can also be found in pure list situations, when the materials were read silently as well. Importantly, I have shown that both production and choice are viable strategies for improving the outcomes of studying. Moreover, unlike production, choice seems to impart only benefits without a known cost. For the studier, this means that there are simple and effective methods of studying, methods without many restrictions on when they can be used, and that do not require more investment than they may be worth. Both production and choice, techniques that anyone can easily do, can aid memory while studying.

## References


Barlow, M. C. (1928). The role of articulation in memorizing. *Journal of Experimental Psychology*, *11*(4), 306-312.

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

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 553-563.

Berry, T., Cook, L., Hill, N., & Stevens, K. (2010). An exploratory analysis of textbook usage and study habits: Misperceptions and barriers to success. *College Teaching*, *59*(1), 31-39.

Bodner, G. E., Jamieson, R. K., Cormack, D. T., McDonald, D. L., & Bernstein, D. M. (2016). The production effect in recognition memory: Weakening strength can strengthen distinctiveness. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(2), 93-98.

Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1711-1719.

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review*, *21*(1), 149-154.

Brown, P. C., Roediger, H. L., III, & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Cambridge, MA: The Belknap Press.

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*(1), 28-35.

Chan, J. C., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*(4), 553-571.

Cho, K. W., & Feldman, L. B. (2013). Production and accent affect memory. *The Mental Lexicon*, *8*(3), 295-319.

Cho, K. W., & Feldman, L. B. (2016). When repeating aloud enhances episodic memory for spoken words: Interactions between production-and perception-derived variability. *Journal of Cognitive Psychology*, *28*(6), 673-683.

Cloutier, J., & Macrae, C. N. (2008). The feeling of choosing: Self-involvement and the cognitive status of things past. *Consciousness and Cognition*, *17*(1), 125-135.

Clump, M. A., Bauer, H., & Bradley, C. (2004). The extent to which psychology students read textbooks: A multiple class analysis of reading across the psychology curriculum. *Journal of Instructional Psychology*, *31*(3), 227-232.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, *22*(1), 267-281.

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*(3), 341-361.

Coverdale, M. E., & Nairne, J. S. (2019). The mnemonic effect of choice. *Psychonomic Bulletin & Review*, *26*(4), 1310-1316.

Cunningham, S. J., Brady-Van den Bos, M., & Turk, D. J. (2011). Exploring the effects of ownership and choice on self-memory biases. *Memory*, *19*(5), 449-461.

Dahlen, K., & Caldwell–Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal*, *97*(4), 902-916.

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 Bulletin & Review*, *8*(1), 155-161.

Driver, C. (2013). A two process account of the memory improvement due to choosing. *Proceedings of the Annual Meeting of the Cognitive Science Society 35*(35), 2219-2224.

DuBrow, S., Eberts, E. A., & Murty, V. P. (2019). A common mechanism underlying choice's influence on preference and memory. *Psychonomic Bulletin & Review*, *26*(6), 1958-1966.

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

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*(7), 1090-1111.

Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (2014). Production improves memory equivalently following elaborative vs non-elaborative processing. *Memory*, *22*(5), 470-480.

Forrin, N. D., & MacLeod, C. M. (2016). Order information is used to guide recall of long lists: Further evidence for the item-order account. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(2), 125-138.

Forrin, N. D., & MacLeod, C. M. (2018). This time it's personal: The memory benefit of hearing oneself. *Memory*, *26*(4), 574-579.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*(7), 1046-1055.

Gates, A. I. (1917). Recitation as a factor in memorization. *Archives of Psychology*, *6*(40), 112-121.

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*(2), 110-119.

Hartley, J., Bartlett, S., & Branthwaite, A. (1980). Underlining can make a difference—sometimes. *The Journal of Educational Research*, *73*(4), 218-224.

Hassall, C. D., Quinlan, C. K., Turk, D. J., Taylor, T. L., & Krigolson, O. E. (2016). A preliminary investigation into the neural basis of the production effect. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(2), 139-146.

Hirano, T., & Ukita, J. (2003). Choosing words at the study phase: The self-choice effect on memory from the viewpoint of connective processing[1]. *Japanese Psychological Research*, *45*(1), 38-49.

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*(4), 534-537.

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/Revue canadienne de psychologie expérimentale*, *62*(4), 242-246.

Hourihan, K. L., & Smith, A. R. (2016). Production does not improve memory for face–name associations. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(2), 147-153.

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). Oxford, UK: Oxford University Press.

Icht, M., Ben-David, N., & Mama, Y. (in press). Using vocal production to improve long-term verbal memory in adults with intellectual disability. *Behavior Modification*.

Icht, M., Mama, Y., & Algom, D. (2014). The production effect in memory: Multiple species of distinctiveness. *Frontiers in Psychology, 5*(886), 1-7.

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/Revue canadienne de psychologie expérimentale*, *70*(2), 154-164.

Jamieson, R. K., & Spear, J. (2014). The offline production effect. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *68*(1), 20-28.

Jones, A. C., & Pyc, M. A. (2014). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 300-305.

Jonker, T. R., Levene, M., & MacLeod, C. M. (2014). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 441-448.

Jonker, T. R., & MacLeod, C. M. (2015). Disruption of relational processing underlies poor memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 831-840.

Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*(1), 121-148.

Kline, C. S. (2019). Production effect in complex texts and over time. *Journal of Integrative Behavioral Science*, *1*(1), 1-7.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 609-622.

Krishnan, S., Watkins, K. E., & Bishop, D. V. (2017). The effect of recall, reproduction, and restudy on word learning: A pre-registered study. *BMC Psychology*, *5*(28), 1-14.

Lambert, A. M., Bodner, G. E., & Taikh, A. (2016). The production effect in long-list recall: In no particular order? *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(2), 165-176.

Lin, O. Y., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *66*(3), 212-216.

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*(2), 237-255.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica*, *98*(2-3), 291-310.

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*(6), 1197-1202.

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*(4), 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*(3), 671-685.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159-163.

Monty, R. A., Geller, E. S., Savage, R. E., & Perlmuter, L. C. (1979). The freedom to choose is not always so choice. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(2), 170-178.

Monty, R. A., & Perlmuter, L. C. (1972). The role of choice in learning as a function of meaning and between- and within-subjects designs. *Journal of Experimental Psychology*, *94*(2), 235-238.

Monty, R. A., Perlmuter, L. C., Libon, D., & Bennet, T. (1982). More on contextual effects on learning and memory. *Bulletin of the Psychonomic Society*, *20*(6), 293-296.

Monty, R. A., Rosenberger, M. A., & Perlmuter, L. C. (1973). Amount of locus of choice as sources of motivation in paired-associate learning. *Journal of Experimental Psychology*, *97*(1), 16-21.

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The simple act of choosing influences declarative memory. *Journal of Neuroscience*, *35*(16), 6255-6264.

Ozubko, J. D., Bamburoski, L. D., Carlin, K., & Fawcett, J. M. (2020). Distinctive encodings and the production effect: Failure to retrieve distinctive encodings decreases recollection of silent items. *Memory*, *28*(2), 237-260.

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*(3), 326-338.

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*(6), 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*(5), 509-524.

Paivio, A. (1971). Imagery and language. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches* (pp. 7–32). New York: Academic Press.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203.

Perlmuter, L. C., & Monty, R. A. (1982). Contextual effects on learning and memory. *Bulletin of the Psychonomic Society*, *20*(6), 290-292.

Perlmuter, L., Monty, R. A., & Kimble, G. A. (1971). Effect of choice on paired-associate learning. *Journal of Experimental Psychology*, *91*(1), 47-53.

Perlmuter, L. C., Scharff, K., Karsh, R., & Monty, R. A. (1980). Perceived control. *Motivation and Emotion*, *4*(1), 35-45.

Putnam, A. L., Ozubko, J. D., MacLeod, C. M., & Roediger, H. L. (2014). The production effect in paired-associate learning: Benefits for item and associative information. *Memory & Cognition*, *42*(3), 409-420.

Quantz, J. O. (1897). Problems in the psychology of reading. *The Psychological Review: Monograph Supplements*, *2*(1), i-51.

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*(8), 904-915.

Quinlan, C. K., & Taylor, T. L. (2019). Mechanisms underlying the production effect for singing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *73*(4), 254–264.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013). How does using object names influence visual recognition memory? *Journal of Memory and Language*, *68*(1), 10-25.

Rommers, J., Dell, G. S., & Benjamin, A. S. (2020). Word predictability blurs the lines between production and comprehension: Evidence from the production effect in memory. *Cognition*, *198*, 1-8.

Sikorski, J. F., Rich, K., Saville, B. K., Buskist, W., Drogan, O., Davis, S. F., ... & Geller, E. S. (2002). Faculty forum. *Teaching of Psychology*, *29*(4), 312-320.

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

Sundqvist, M. L., Mäntylä, T., & Jönsson, F. U. (2017). Assessing boundary conditions of the testing effect: On the relative efficacy of covert vs. overt retrieval. *Frontiers in Psychology*, *8*(1018), 1-15.

Taikh, A., & Bodner, G. E. (2016). Evaluating the basis of the between-group production effect in recognition. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(2), 186-194.

Takahashi, M. (1989). The self-choice effect in learning and memory. *Kyoto University Research Studies in Education*, *35*, 211-221.

Takahashi, M. (1992). Memorial consequences of choosing nonwords. *Japanese Psychological Research*, *34*(1), 35-38.

Takahashi, M. (1997). Kioku ni okeru fugoka horyaku no kenkyu [Studies of encoding strategies in memory]. Kyoto, Kitaoji-shobo.

Toyota, H. (2015). The role of word choice and criterion on intentional memory. *Perceptual and Motor Skills*, *120*(1), 84-94.

Toyota, H. (2016). The effects of choosing method to use for memorization on intentional memory of words. *奈良教育大学紀要. 人文・社会科学, 65*(1), 21-26.

Toyota, H., & Tsujimura, M. (2000). The self-choice elaboration effects on incidental memory of Japanese historical facts. *Perceptual and Motor Skills*, *91*(1), 69-78.

Voss, J. L., Gonsalves, B. D., Federmeier, K. D., Tranel, D., & Cohen, N. J. (2011). Hippocampal brain-network coordination during volitional exploratory behavior enhances learning. *Nature Neuroscience*, *14*(1), 115-120.

Watanabe, T. (2001). Effects of constrained choice on memory: The extension of the multiple-cue hypothesis to the self-choice effect. *Japanese Psychological Research*, *43*(2), 98-103.

Watanabe, T., & Soraci, S. A. (2004). The self-choice effect from a multiple-cue perspective. *Psychonomic Bulletin & Review*, *11*(1), 168-172.

**Appendix A**

# Toucans

Keel Billed Toucan – the most colorful toucan

Toco Toucan – the largest toucan

## The Basics

**E**veryone who has seen a Toucan has surely noticed its huge bill. Scientifically the Toucan family is called Ramphastidae. Toucans get their name from 'tucano' given to them by the Tupi Indians of Brazil. The average lifespan of a well-cared for toucan is at least ten years, and they can live to be approximately 15 years old. The toucans are poor fliers because of the size of their bills, and they rely on hopping from branch to branch in trees.

It's striking that the relation between the size of the Toucan's bill and the size of its body increases with the size of the species. Toco toucans (the largest toucan) have the largest bill of any toucan. The insides of their bills are shaped like honeycombs. It is hollow, made of protein keratin with thin rods of bone to support it - similar in

consistency to a hard sponge. A thin outer sheath encloses a hollow which is crisscrossed by many thin, bony, supporting rods.

The toucan's large bill can be used for many purposes. It's especially useful when foraging. Their bills enable them to perch inside the crown of a tree, where branches are thicker, and reach far outwards to pluck berries or seed from twigs too thin to bear their weight. Seized in the tip of the bill, food is thrown back into the throat by an upward toss of the head. When the fruits are too large, toucans use their bills (which have serrated edges) to tear these fruits into smaller pieces. Fruits dominate their diet, but Toucans are not pure vegetarians; they also hunt on insects and small reptiles - another field of activity for the large bill. It also plays an important part concerning communication, especially during courtship. In part thanks to their bill Toucans are well-fortified birds who are able to defend their young against some predators. The varied patterns of toucans' bills may also help these birds to recognize each other or attract a mate. During their nuptial display, both partners play a game in which they throw berries to each other or toss them against each other with their bills.

But there is one thing they can't use the large bill for: the construction of tree holes. Toucans need such holes for nesting and raising their young. Although they are closely related to woodpeckers, they are not able to construct such holes. Toucans depend on natural tree holes or holes constructed by other animals. When toucans sleep, they turn their head so that their long bill rests on their back and their tail is folded over their head. The bird becomes a ball of feathers. Often found in abandoned tree hollows or old woodpecker holes, five or six adults may sleep in one hole!

Parents share incubation and are, for birds of their size, impatient sitters. They rarely remain at their task for more than an hour and often leaving their eggs uncovered. The female lays 2 to 4 eggs, which are incubated by both parents. The young Toucans hatch after an incubation period of a little more than two weeks. They grow very slowly. After about 16 days of incubation the nestlings hatch blind and naked, still with no trace of feather on their pink skins. Nestlings are fed by both parents, with increasing quantities of fruit. The feathers of the small toucanets do not begin to expand until they are nearly four weeks old. Parents will care for their young for about 8 weeks; however, small toucanets may fly from the nest when they are 43 days old. When they are fully-fledged, they will still come back to their breeding hole for quite a while. The parents and their young stay together for several months until the family group breaks off.

Toucans, as well as other birds of the same species, are indigenous to tropical America. By nature, toucans are rain forest dwellers, and live high among treetops. In its native region, toucans are associated with evil spirits and are thought to be the incarnation of a demon. In certain religions of South and Central America, the father of a new child must not eat toucan flesh as it might bewitch the newborn and cause it to fade away. The toucan can also be a tribal totem and the medicine man can use it as an incarnation to fly to the spirit world.

## Rearing Toucans

**B**efore the Wild Bird Conservation Act of 1992 was passed toucans were exported from the Mexican, Central, and South American neotropical regions and shipped to the US and many other countries of the world.

During the late 70's and 80's more and more toucan species became harder to acquire and there were already fewer breeders in the US who kept them in their collections. Parrots were more popular than many softbills because they were easier to keep and they lived longer. Therefore, toucans were imported less frequently and trappers started devoting their time to capture parrots. Breeders also concentrated on parrots to keep up with the demand. Parrots began appearing on popular TV shows (notably Fred the sulphur-crested cockatoo on the 1975-78 series "Baretta"), which contributed to the boom. Meanwhile, toucans were slowly disappearing from America and if not for the few who continued the struggle with their health problems and making these birds their passion, all of them would have vanished from American aviculture. Once common in American aviculture, the toucan is now few in numbers.

The Defenders Of Wildlife, an organization dedicated to the protection of all wild animals and plants in their natural communities, in conjunction with other conservation and humane organizations, ornithologists, bird breeders and the pet industry spearheaded development of the Wild Bird Conservation Act (WBCA). Passed into law by the United States Congress in 1992, the WBCA restricted or halted the importation of all bird species listed in the Appendices of the Convention on International Trade in Threatened and Endangered Species (CITES). This ended commercial import to the US of many bird species including toucans. There are currently 5 toucan species listed on CITES and it is now illegal to import them. Before the WBCA was enacted, the US was the world's largest buyer of wild birds, importing more than 7.4 million birds between 1980 and 1991. The number of birds imported dropped significantly after the enactment of the WBCA. Throughout the 1980s, the US imported an average of 700,000 birds annually, according to the U.S. Department of Agriculture.

With the WBCA now in effect and stricter new control of importation of non-native animals into the US, the toucan became a rare and difficult bird to acquire. During the 1990's and into the millennium many people thought owning a toucan was at least unheard of or even illegal. When you mention a toucan to somebody today their first or second thought would almost certainly be of that colorful "Toucan Sam" on the cereal box.

## Diet and Hemochromatosis

**O**ne of the biggest challenges of keeping and maintaining a large softbill collection or just a single pet is the problem of the iron storage disease known as hemochromatosis. This disease affects a wide range of species including toucans, mynahs, tanagers, birds-of-paradise, and many other softbills. It is not a disease we see in parrots, finches, or canaries. Hemochromatosis is probably the leading cause of death of mynas and toucans in captivity, and collections of birds-of-paradise have been decimated by this malady.

The cause is unknown, but the effect is a massive uptake and storage of iron in the liver. This excessive storage leads to hepatomegaly or liver enlargement. Consequently, some degree of liver dysfunction is observed. However, the most remarkable aspect of the pathology of the disease is liver enlargement. This enlargement leads to ascites or fluid exuded into the air sacs. This fluid in combination with an oversized liver results in difficulty in breathing. The clinical presentation of these birds is often a myna or toucan with a swollen abdominal area accompanied by open-mouthed breathing. The problem can be so severe in some mynahs that if you changed their position from upright to lateral, the birds began to drown in their own fluids, which were leaking from the air sac into the lung. Large quantities of fluid may be taken from a small bird.

If the bird is at adequate surgical risk, a liver biopsy can be taken for histopathological examination. The pathologist can confirm the diagnosis by typical golden-brown deposits of iron in the majority of liver cells using specialized staining techniques. The clinician can also make a tentative diagnosis by examining certain liver enzymes. Often, an X-ray is helpful in determining liver enlargement but does not confirm a positive diagnosis without proper histopathological results.

Once the diagnosis is reached, treatment options are very limited. Hemochromatosis also occurs in humans, who can be treated with the drug penicillamine. Unfortunately, such treatments do not seem to work in birds, probably because the causes are different. Drugs like the aforementioned penicillamine, which control human cases, are ineffective in birds. However, a viable treatment, although ancient in origin, is phlebotomy (or bleeding). The logic behind this procedure is: (a) blood contains iron, (b) the liver contains excess iron, (c) phlebotomy causes anemia and more iron is needed to make more blood, and (d) the liver provides the source of iron so the excess iron is depleted via natural mechanisms. This therapy has been used with some success in toucans. But phlebotomy must be repeated on a timely basis, usually once or twice a month, and must be continual. Obviously, this is simply a palliative measure because the inciting cause, which is unknown, has not been removed. Also, if the heart is involved, therapies useful in treating cardiac deficiencies must also be used.

The obvious answer is prevention. Current knowledge strongly suggests that the cause is dietary, although other theories such as a viral cause have been advanced. Several commercial bird food manufacturers market low iron foods for toucans and other softbills. These foods, although of high quality, have been used with varying degrees of success. But they are the best alternative we have now.

Some noted softbill aviculturists use a cheap quality dog food and have had little problem with respect to hemochromatosis. There is a possibility that this dog food is a soya based product and is supplemented with inorganic iron, a cheap way to make dog food. However, this provides a vegetable-based protein for the birds and perhaps of a lower iron bioavailability. Supplying your toucans with fresh fruit keeps them happy and hydrated. Toucans in the wild survive on a nearly pure fruit diet, albeit different fruits than we supply in captivity. In addition to fruits, nutrient pellets should be provided as a "side dish" to supply protein, vitamins, and minerals that our fruits do not provide.

The bottom line is that if you deal with softbills, you will encounter hemochromatosis sooner or later. A diet low in iron, especially low in highly bioavailable iron sources, should be given to the birds until the cause of hemochromatosis can be established.

**Appendix B**

Set A Questions

The largest toucan species is the _____.

The insides of toucan's bills are shaped like _____.

What is the usage of toucans' bills during nuptial display? _____

What other bird species is the toucan related to? _____

Small toucanets may fly from the nest when they are _____ days old.

Toucans mainly inhabit the _____ in South America.

In its native region, toucans are associated with evil spirits. The father of a new child must not _____ as it might bewitch the newborn.

The Defenders of Wildlife, in conjunction with other organizations, spearheaded development of the _____ (*you may use abbreviation*).

The iron storage disease that killed many toucans is called _____.

One method helps, but does NOT confirm, the positive diagnosis of iron storage disease. What is it? _____

If an iron storage disease is diagnosed in human, the primary drug to use is _____.

One can feed toucans with cheap dog food if the dog food is made with _____.

Set A Answers

toco toucan

honeycombs

toss berries between partners

woodpeckers

43

rain forests

eat toucan flesh

WBCA

hemochromatosis

Using X-ray to examine liver enlargement

penicillamine

soya-based product

Set B Questions

The most colorful toucan species is the _____.

The insides of the toucan's bills have a similar consistency to _____.

What is the usage of a toucan's bill when they encounter fruits that are too large to swallow? _____

Toucans sleep inside _____.

After ___ days of incubation the nestling are still featherless.

The primary way for toucans to travel is _____.

In its native region, the toucan can be a tribal totem. The medicine man can use it as an incarnation to _____.

The United States Congress, since 1992, restricted the importation of all bird species listed in the Appendices of _____. (*you may use abbreviation*)

The iron storage disease that killed many toucans is not seen in birds like _____. (give 1 example)

One way to tell that a toucan is suffering from iron storage disease from physical appearance alone is _____.

What is the primary treatment if an iron storage disease is diagnosed in a toucan? _____

In addition to fruits, one should feed toucans with _____.

Set B Answers

keel billed toucan

a hard sponge

they tear them into smaller pieces with their bill.

tree holes

16

hop among treetops

fly to the spirit world

CITES

parrots, finches, or canaries

swollen abdominal area and open-mouthed breathing

phlebotomy

nutrient pellets