

Reading aloud as a distinctive context:

Examination of the production effect as a context-based memory effect

by

Yichu Zhou

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, 2022

© Yichu Zhou 2022

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

Tracy Taylor-Helmick

Professor, Department of Psychology and
Neuroscience, Dalhousie University

Supervisor

Colin M. MacLeod

Professor, Department of Psychology

Internal Members

Myra A. Fernandes

Professor, Department of Psychology

Evan F. Risko

Associate Professor, Department of Psychology

Internal-External Member

Steve E. Mock

Associate Professor, Department of Recreation and
Leisure Studies

Author's Declaration

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

Research on the production effect has established that studying information aloud is a simple memory technique useful for learning new information in various settings; this memory benefit has been credited to the distinctive processing of information studied aloud relative to other information studied silently. In this dissertation, I aim to characterize the production effect more broadly as a context-based memory effect, thus expanding how production is conceptualized as a memory technique. Essentially, at encoding, the distinctive “aloud” information creates a global contextual cue which becomes associated with (only) the produced information, and this cue can then be elicited at retrieval to facilitate memory for the produced information. This provides a more concrete mechanism than is provided by the current conceptualization that information studied aloud (or produced in another way) benefits only by being distinct from other information studied silently: Distinctiveness is underpinned by context.

Three studies are reported, the first two using recognition testing. In Study 1 (Chapter 2), a pure-list production effect manipulation—where participant study only aloud items or only silent items—was inserted into a mixed-list design containing both aloud and silent items; the goal was to examine whether the co-occurring mixed-list procedure would enhance the magnitude of the pure-list production effect by prompting participants to use the distinctive information—production—to support also the encoding and retrieval of the pure-list aloud items. A larger-than-typical pure-list production effect was indeed found: Presumably, production was encoded as a contextual cue associated with the pure-list aloud items, which subsequently aided the retrieval of these items because the non-studied items on the recognition test were associated with a different context. Thus, aloud items studied in pure lists can be sufficiently differentiated without requiring a cost-benefit trade-off as in mixed-list designs, where some silent information

must suffer a memory cost at the same time that aloud learning is enhanced. In addition, the secondary findings suggested that contextual overloading may in fact underlie the previously reported influence of statistical distinctiveness within the production effect, bolstering the argument for the production effect to be considered as a context-based memory effect.

In Study 2 (Chapter 3), a mixed-list production effect manipulation was inserted into a list-method directed forgetting procedure to investigate whether attempting to use production as a memory cue would aid the later retrieval of forgotten aloud information. According to the contextual change hypothesis of list-method directed forgetting, induced forgetting reduces memory of a list of items due to a mental context change, and reinstating the relevant contextual cues associated with the forgotten list will in turn improve memory of that list; the aim of the manipulation was to test whether production belongs to this conceptualization of contextual cues. The results reinforced that reading aloud does indeed function as contextual information in the production effect, in that reactivating the use of this information at retrieval enhanced memory of only aloud items—and not silent items—intentionally forgotten.

Study 3 (Chapter 4) explored whether production, as contextual information, also supports the memory of information studied aloud when the method of retrieval is recall. A modified mixed-list procedure was used where pairs of aloud and silent lists were studied in a series of trials, and performance was compared to that on single-list trials—consisting of only one aloud or one silent list—to gauge whether a produced list, due to its distinctive context, would be more protected from interference effects of studying another list concurrently compared to an unproduced list. The primary finding was decreased proactive interference from earlier information when newer information is studied aloud: This illustrates a potential release

from proactive interference by means of a context change when the more recent information is studied aloud, regardless of how the earlier information was studied.

Taken together, these studies show that the production effect is a useful learning technique that enhances memory through distinctive context-based effects.

Acknowledgements

To my supervisor, Dr. Colin M. MacLeod, I am extremely grateful to have received his expertise and guidance throughout the past few years which has made the completion of the present work possible. His support has undoubtedly made the experience more enjoyable and less stressful, particularly in the face of the COVID-19 pandemic which began near the completion of my program. I also greatly appreciate the members of my Examining Committee for taking the time to provide invaluable feedback to my work.

Thank you to Dr. Roxane J. Itier and the members of the Face Processing and Social Cognition Lab for their willingness to provide the much-needed expertise and guidance during our EEG project collaboration. Although we were unable to complete the project due to COVID-19, I have learned an immense amount from the experience which would not have been possible without the support of Dr. Itier and her lab.

I am thankful for my labmates in the Memory, Attention, and Cognition Lab, as well as all of the members in the Cognitive Area of the Department of Psychology, for always being willing to provide insightful feedback and advice to my work. Thank you also to the research assistants who have supported my work and enabled its completion.

Lastly, thank you to my family and friends for their love and encouragement throughout my PhD program.

Table of Contents

List of Figures	x
List of Tables	xi
Chapter 1 – Introduction	1
The Production Effect	2
The Distinctiveness Account and Recognition.....	3
The Item Order Account and Recall.....	6
A Context-Based Perspective on the Production Effect.....	8
Rationale for the Experiments.....	12
Chapter 2 – The Pure-List Production Effect and Statistical Distinctiveness	18
Experiment 1	22
Method.....	23
Results	24
Experiment 2	26
Method.....	27
Results	28
Experiment 3	32
Method.....	32
Results	33
Discussion	36
Chapter 3 – Examining the Production Effect using Directed Forgetting	47
Method	54
Results.....	58

Discussion	68
Chapter 4 – The Production Effect in Recall	76
Method	82
Results	86
Discussion	92
Chapter 5 – General Discussion.....	102
Summary of the Findings	103
Theoretical Perspectives: Production and Context.....	108
Limitations and Future Directions.....	114
Conclusion.....	117
References	119
Appendix A.....	135
Appendix B	137
Appendix C	139

List of Figures

Figure 1. Proportions of hits for studied targets in Chapter 2.....	26
Figure 2. The study procedure of Chapter 3	57
Figure 3. Proportions of hits for studied targets in Chapter 3.....	60
Figure 4. Proportions of studied targets that were correctly recognized but that were associated with the incorrect source in Chapter 3	66
Figure 5. Proportion correct recall in Chapter 4	88

List of Tables

Table 1. Proportions of hits and false alarms in Chapter 2	25
Table 2. Proportions of hits and false alarms in Chapter 3	59
Table 3. Proportions of studied targets that were correctly recognized but that were associated with the incorrect source in Chapter 3	65
Table 4. The ten study trials in Chapter 4	77
Table 5. Proportion correct recall and average number of intrusions for each of the 10 experimental trials in Chapter 4	87

Chapter 1 – Introduction¹

There have been few encoding techniques that research has consistently demonstrated to deliver memory benefits when the opportunity for studying information is limited. The oldest is imagery (Paivio, 1971): Visually picturing the information that needs to be learned is an encoding method that reliably enhances memory. Generation (Slamecka & Graf, 1978) is also a well-established technique, where information generated by the learner from given cues is better remembered than if the same information is simply provided to them. Another such technique is enactment (Cohen, 1981)—performing the actions described by a phrase leads to better memory than simply reading that phrase. Recently, the drawing effect (Fernandes et al., 2018)—the benefit from drawing pictures rather than writing, has joined this exclusive group.

All of these techniques have in common the limitation that they cannot be performed simply by the learner. It is not possible in all situations to enact or even to imagine the information that needs to be remembered, and generation and drawing require time and effort from the learner in creating cues or drawings associated with the target information during encoding (although even preparing to draw appears to confer a memory benefit: see Wammes et al., 2018). In the past decade, an encoding technique called the *production effect* (MacLeod et al., 2010) has gained considerable interest in the memory research literature as a very simple-to-use mnemonic compared to the other listed techniques, a technique which nonetheless enhances memory just as reliably.

The production effect—a memory benefit for produced versus unproduced information (typically referring to reading information aloud versus silently)—is a robust effect that has been

¹ Sections of this chapter were taken from Zhou and MacLeod (2021) published in *Memory*, slightly modified from the published version to fit the structure of this dissertation.

demonstrated across numerous studies (see the special issue of the *Canadian Journal of Experimental Psychology*: Bodner & MacLeod, 2016; see MacLeod & Bodner, 2017, for a brief review). The goal of this dissertation is to demonstrate that the production effect can be generalized as a context-based memory effect, which will be demonstrated across three converging manipulations. In this chapter, I begin by providing an overview of the relevant literature on the production effect and on context-based memory effects. Then, I will present a summary of the rationale for the three studies constituting this dissertation, to be followed by more specific rationales for each of the three studies in Chapters 2 through 4. Chapter 5 will summarize the empirical work in the dissertation and focus on the theoretical account.

The Production Effect

Prior to being named and brought into the spotlight by MacLeod et al. (2010), a memory benefit from producing information had already been demonstrated in several earlier studies, although such research had been sparse. One example was a study by Barlow (1928), who showed that studying consonant-vowel-consonant nonwords by reading them aloud enhanced subsequent recall compared to studying them silently. The beginning of the modern production effect literature is usually attributed to Hopkins and Edwards (1972), who demonstrated a production effect in a mixed-list design where some words were studied aloud whereas other words were studied silently; in a pure-list design, however, where separate groups of participants studied the words either all aloud or all silently, they observed no production effect.

Other studies had also reported production effects (Conway & Gathercole, 1987; Dodson & Schacter, 2001; Gathercole & Conway, 1988; Kurtz & Hovland, 1953; MacDonald & MacLeod, 1998), but they all had received relatively scant attention until the phenomenon re-emerged in the past decade. In the current era, the production effect has been shown to extend

beyond speaking words: Writing (Forrin et al., 2012), typing (Forrin et al., 2012; Jamieson & Spear, 2014), and mouthing (MacLeod et al., 2010) all result in a production advantage, although none of these enhances memory as much as speaking does. In addition, the production effect has been generalised to studying educationally relevant material like text and to longer retention intervals (Ozubko, Hourihan, & MacLeod, 2012), as well as to learning new vocabulary in a second language (Icht & Mama, 2019), indicating that production can be effective as a learning strategy beyond the laboratory. MacLeod and Bodner (2017) present a brief review of the findings and their interpretation.

The Distinctiveness Account and Recognition

The production effect observed in recognition memory has been explained primarily as due to the distinctive processing applied to the produced words (Conway & Gathercole, 1987; Forrin et al., 2014; MacLeod et al., 2010; Quinlan & Taylor, 2013). Under the distinctiveness account, it is hypothesised that the aloud words stand out relative to the silent words during study, resulting in an additional dimension of encoding for the produced words. By analogy, the aloud words are seen as figure whereas the silent words are seen as ground. Hunt (2013, p. 10) defines distinctive processing as “the processing of difference in the context of similarity”. Thus, the words are all similar in a variety of ways, including being common nouns and being presented in lowercase font at the centre of the screen. But the aloud words differ from the silent ones by virtue of being produced, and that makes their encoding distinctive.

The importance of the distinctive processing during study becomes apparent at the time of test. In addition to trying to remember the word itself, participants can use the strategy—consciously or unconsciously—of trying to retrieve whether the word was produced at study. On a recognition test, then, remembering having spoken a word during the study phase is an

additional way to certify that it was indeed studied. Therefore, the distinctiveness dimension provides an additional path for successful recognition, augmenting memory for the word itself. Dodson and Schacter (2001) characterised this type of retrieval strategy as a distinctiveness heuristic.

Evidence for the distinctiveness account of the production effect comes from two main types of results. First, the production effect typically is much larger in within-subject mixed list designs, in which a participant studies both aloud and silent words, than in between-subjects pure list designs, in which a participant experiences only one of these conditions (Hopkins & Edwards, 1972; MacLeod et al., 2010; see Fawcett, 2013). This makes sense in that establishing the contrast between aloud and silent items during study is possible only under the within-subject design. Second, Fawcett and Ozubko (2016) used remember-know judgments as well as a signal detection approach to further examine the processes underlying the influence of this design difference. Under a within-subject design, they observed a production effect both for the words that participants “recollected” and for the words that participants indicated were just “familiar”, essentially replicating an earlier study by Ozubko, Gopie, and MacLeod (2012). In contrast, under a between-subjects design, they observed a production effect only for familiarity—that is, the between-subjects effect did not involve a recollective-based component.

These findings suggest that experiencing both aloud words and silent words at the time of study plays a critical role in making the aloud words distinctive, thereby producing an enhanced memory benefit via recollection—in particular, recollection of having said a word aloud. In support of this idea, studies have shown that participants’ source memory for whether a word was produced at study is better for words that were studied aloud than for those that were studied silently (Ozubko, Gopie, & MacLeod, 2012) and is independent of manipulated memory strength

of the studied items (Ozubko et al., 2014). Older adults, who do not use distinctive information as well as younger adults, also show a smaller production effect (Lin & MacLeod, 2012). These findings support the idea that, in addition to encoding the words themselves, participants also encode whether they studied them aloud; that information can then be retrieved at test to assist remembering.

Formal models have also demonstrated that the memory advantage for produced information can be accounted for by the encoding of additional distinctive features for aloud versus silent items during study. Jamieson et al. (2016) used MINERVA2 as their basis, a model which aims to account for a wide range of memory phenomena by supposing that information is encoded as episodic traces (representing each time the information is encountered), each with unique associated features. For the production effect, encoding of produced information was presumed to include additional sensory features (i.e., features which are related to the act of production) compared to unproduced information; these additional features are then utilized as a memory cue during later retrieval of the produced information—as hypothesized by the distinctiveness account. Their adapted model was able to predict production effects observed in both mixed-list and pure-list designs, as well as a relatively smaller production effect in pure-list designs. Using a Retrieving Effectively from Memory (REM) framework, Kelly et al. (in press) proposed a similar mechanism wherein the memory benefit for produced information could be modelled by incorporating the use of additional production-related features. Critically, their modelling also showed that the use of these additional features at retrieval required additional time: The magnitude of the production effect was reduced as retrieval time became limited, indicating a decreasing effectiveness of the production-related features as memory cues for

retrieval (see also Chapter 3 for a discussion of the use of contextual features/cues at retrieval as a time-dependent process).

Although it has been widely reported that the production effect is considerably smaller in between-subjects, pure list designs than in within-subject, mixed list designs—indeed, the effect initially was thought to be restricted to within-subject designs (Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010), a belief that formed the original foundation for the distinctiveness account—it is now clear that the effect is present under both designs, albeit smaller between-subjects (see MacLeod & Bodner, 2017, for a brief review). The existence of a pure-list production effect points to a possible strength-based contribution to the effect, in which memory for produced information is enhanced compared to unproduced information due to an increase of memory strength provided by the act of production (Bodner & Taikh, 2012; Jamieson et al., 2016). It remains possible, however, that a distinctiveness explanation could also account for the pure-list production effect: On a recognition test, the non-studied distractors can also be contrasted with the aloud items—because these distractors had not been produced. The distinctiveness heuristic could thus be used as well for the retrieval of aloud items in pure-list designs, although this strategy would understandably be less apparent for participants when all of the information was studied aloud; pure-list production effects may therefore be relatively smaller because the distinctiveness heuristic is inconsistently and less routinely applied in these situations.

The Item Order Account and Recall

Thus far, I have been describing studies that have used recognition testing. Distinctiveness—albeit in a broader sense—has also been put forth as an explanation of the production effect shown on recall tests; however, the specific cognitive mechanism that underlies

distinctive processing in the case of recall is less clear. Recent studies examining the production effect using recall tests (e.g., Forrin & MacLeod, 2016; Lambert et al., 2016; Jones & Pyc, 2014; Jonker et al., 2014) have provided evidence in support of an influence of item-order memory on the production effect observed in these situations. Relative to studying pure lists consisting of items studied either all aloud or all silently between subjects, where recall has consistently failed to reveal a production effect, studying a mixed list containing both aloud and silent items within-subject has typically resulted in only a minimal memory benefit for the aloud items whereas the silent items appear to suffer a memory cost.

According to an item-order account of the production effect in recall (Jonker et al., 2014), memory for the silent items in mixed-list designs suffers due to the more distinct (“unusual”), item-specific, encoding process of the aloud items, which captures attention and disrupts relational processing of the silent items that otherwise would act as a key mechanism for retaining information studied in a relatively common manner. The use of this order memory technique is specific to the examination of recall where, unlike on recognition tests, the order of retrieving the items is under the control of the participant. In a pure-list design, distinctive processing benefits the encoding of aloud items whereas relational processing benefits the encoding of silent items. In a mixed-list design, however, only the benefit of distinctive processing for the aloud items is maintained because the aloud and silent items are presented in random order which defeats relational processing, resulting in a memory decrement for the silent items relative to a pure-list design.

The item-order account originated in the context of the generation effect (Slamecka & Graf, 1978), in which words produced from hints are better remembered than words that are simply read. This is a substantial effect in within-subject mixed list experiments. Nairne et al.

(1991) showed that when participants studied pure lists either by generating all of the words or by simply reading all of the words, there was no benefit of generation and there was poorer memory for the serial order of the generated lists relative to the read lists. On this basis, Nairne et al. (1991) proposed that in pure lists the memory benefit from generation that is routinely seen in mixed lists—due to the distinctive encoding of the generated words—was offset by the relatively superior encoding of relational—in this case, order—information within the read lists (which represented the relatively common encoding technique in their study).

Subsequently, McDaniel and Bugg (2008) extended the item-order account to other learning tasks that involve the distinctive processing of some information during encoding, an example being enactment (Cohen, 1981), where the actions associated with words are physically performed. Additionally, McDaniel and Bugg proposed that whereas within pure lists there is a memory benefit from the encoding of relational information for items studied using a relatively common technique, in mixed lists, order memory for these items is disrupted by any co-occurring unusual encoding processes, resulting in a memory cost for the “common” items compared to when these are studied within pure lists. Results that are consistent with these predictions have been shown in phenomena including generation (Hirshman & Bjork, 1988) and enactment (Engelkamp & Dehn, 2000). The item-order account of the production effect proposed in Jonker et al. (2014) would fall under this broad umbrella of item-order effects: In their study, silent items studied in pure lists indeed showed better retention of serial order compared both to aloud items studied in pure lists and to silent items studied in mixed lists.

A Context-Based Perspective on the Production Effect

The central question motivating the present research pertains to whether the effect of distinctiveness—exhibited in both recognition and recall—can be characterized as a context-

based memory effect. Specifically, when information is produced during study, does this additional dimension of encoding essentially result in a contextual cue being stored with the produced item? Then, at the time of retrieval, can this contextual cue with which the information produced at study is associated be recovered to facilitate remembering of the produced information?

Broadly, we know that context is profoundly influential in remembering, as has long been emphasized in theorizing (e.g., McGeoch, 1932). Contextual similarity between encoding and retrieval has been presented as a key factor in determining whether studied information will be remembered: Such effects have been reported in studies investigating environmental context changes (e.g., Godden & Baddeley, 1975; Isarida & Isarida, 2007; Shin et al., 2021; Smith & Manzano, 2010; Smith et al., 1978; see Smith & Vela, 2001, for a review) as well as changes in physical or mental states (e.g., Eich, 1980; Teasdale & Fogarty, 1979) between study and retrieval. Formal modelling has also incorporated the influence of contextual cueing on memory (see, e.g., Howard & Kahana, 2002; Malmberg et al., 2006; Mensink & Raaijmakers, 1988, 1989), suggesting that, as a memory aid, people encode the active contextual cues at study and then attempt to use these contextual cues to remember the associated studied information at retrieval. Since context presumably changes with time—the idea of contextual drift is prevalent in these formal models—temporal delays in retrieval would thus mean that it becomes more difficult to use the contextual information at retrieval as associative cues for remembering the studied information, in which case forgetting would be observed.

The phenomenon of context-based remembering can be characterized as an effect of encoding specificity (Tulving & Thomson, 1973): Overlap of contextual cues between encoding and retrieval facilitates memory of studied information. Alternatively, it can be characterized as a

facilitation effect from the use of mental schemas: Learning and retrieving information are facilitated when connections are made between new information and existing mental schemas, with schemas acting as scaffolds which help to organize the new information (e.g., Bransford & Johnson, 1972; Tse et al., 2007). By analogy, a contextual cue can also serve as a mental schema to which newly acquired information becomes attached, thereby facilitating memory when the same contextual cue is reinstated at retrieval (Shin et al., 2021).

To date, research has not specifically defined the production effect as a possible context-based memory effect. Yet, as discussed in the previous section, source memory for produced information—as an associative detail—has been shown to contribute to the production effect, at least under recognition testing (Fawcett & Ozubko, 2016; Ozubko, Gopie, & MacLeod, 2012). By demonstrating that the production effect involves a recollective process—defined as being able to mentally re-experience or remember the episodic details of past events (see Yonelinas, 2002, for a review)—we know that remembering that an item was produced is at least partly responsible for the memory enhancement for these produced items (this is, of course, the core proposal of the distinctiveness account). This idea is consistent with previous studies demonstrating that recollection of target information tends to be accompanied by memory of source, or contextual, details associated with that information (e.g., Eldridge et al., 2005; Hicks et al., 2002; Perfect et al., 1996; for reviews, see Johnson et al., 1993; Mitchell & Johnson, 2009)—and certainly also to a greater extent than information remembered through only familiarity (i.e., a vague feeling, compared to recollection, that an event had been experienced). These findings suggest that production could be characterized as a contextual detail which becomes associated with produced items, and memory of this contextual cue would therefore

appear to support memory for the produced items themselves through the linkage of the item information with the source information.

In this dissertation, I thus aim to combine the core theoretical ideas of the production effect and context-based memory effects by proposing that within the scope of the production effect, the act of producing an item—though item-specific—also becomes associated with these produced items as an overarching *global* (e.g., list-wide) contextual framework. In pure-list designs containing only produced information to be encoded, production would simply be a contextual detail to which all of the target information becomes associated—like context-based remembering in the typical sense. However, in the case of mixed-list encoding designs containing both produced and unproduced information, production only becomes selectively associated with the produced target information as a contextual detail (where it remains a global contextual framework to which all produced information becomes associated), because encoding non-produced information obviously does not involve the act of production and also is not distinct relative to produced information.² In both the pure-list and the mixed-list designs, intentionally reinstating the use of this production framework at retrieval should facilitate memory for the items that were produced during study but that are now proving difficult to remember based on item information alone.

To be clear, the hypothesis is that the *specific information that is to be learned* must be produced for this particular context-based memory benefit of self-production to occur. The reason is twofold: (1) production is unique compared to contextual elements that simply exist in the background—it is an action that one takes only at specific times (i.e., when actively encoding

² Hence, in mixed-list designs, the conceptualization of production as a *global* contextual detail stands in contrast to memory tasks where to-be-encoded items are differentially associated with a series of “localized” contextual cues—such as various background pictures (Murnane et al., 1999)—because memory of the non-distinctive unproduced information is not supported by any equivalently distinctive contextual associations.

information); and thus (2) when one is speaking during encoding, producing speech different from the to-be-encoded information would not establish a context-to-item link between production (the contextual cue) and the information needing to be remembered because the learner is distracted from the encoding process. Therefore, the overarching contextual cue of production combines with the item-specific action during encoding to support memory, whereas producing other things apart from the to-be-remembered material—for example, non-specifically saying “yes” to every item that is to be produced instead of speaking the individual words (MacLeod et al., 2010, Experiment 4B)—does not lead to a memory benefit. As will be described in more detail in the next section, the mechanism in essence is predicted to be similar in both recognition and recall, although the observed effect of production on memory appears to be somewhat different under the two retrieval methods (notably in the substantial memory cost for the unproduced information in recall) due to the way in which information is required to be recovered from memory.

Rationale for the Experiments³

As stated earlier, the goal of this dissertation is to demonstrate that the production effect can be generalized as a context-based memory effect. To accomplish this, three converging manipulations will be investigated over the next three chapters. The first two studies, covered in Chapters 2 and 3, explore this context idea in production effects seen in recognition memory. The third study, covered in Chapter 4, examines how this idea may exhibit itself differently when recall is used as the retrieval test instead. The aim for both types of tests is to demonstrate that

³ As this section is only intended to provide an overview of the experiments in this dissertation, detailed references to the relevant literature for each study will be provided in Chapters 2 through 4.

reading information aloud serves as a useful context-based memory technique which is applicable to improving memory for key information.

The first study (Chapter 2) considered the hypothesis that the distinctiveness heuristic is used inconsistently in the encoding of pure lists, and investigated whether, by implicitly showing participants that the distinctiveness heuristic can be useful as a context-based memory technique even when pure lists containing only aloud items or only silent items are studied, a robust production effect can be elicited in the study of pure lists, close to that seen when studying mixed lists. In this study, aloud versus silent processing was manipulated within subject in a mixed-list procedure. Additional pure-list items—read either all aloud or all silently—were alternated with these mixed-list words. Participants were instructed that the mixed-list items would be tested and were falsely led to believe that the pure-list items would not be tested, which acted to differentiate the two types of lists.

The hypothesis was that when the pure-list items were in fact tested, a between-subjects production effect—one that might be rather large—would be observed. This would be attributed to the distinctiveness of the pure-list aloud words being enhanced by the co-occurring within-subject manipulation: Participants would become aware that they could use the distinctiveness heuristic (i.e., remembering whether items were studied aloud) to improve their memory of the pure-list aloud items because of the co-occurring mixed-list items. With regard to the central hypothesis of this dissertation, production would act as the global contextual framework specifying the method of encoding to which the pure-list aloud items would become associated; the non-studied distractor items on the recognition test would be differentiated by virtue of not having been produced. This expected result would support the hypothesis that distinctiveness, as a context-based memory technique of remembering whether information was produced during

encoding, can enhance memory for information studied aloud even in pure-list designs, and that previous studies may have shown smaller production effects in pure-list relative to mixed-list designs due to participants inconsistently and less routinely using the distinctiveness technique because it was not apparently useful in these situations.

Whereas the first study manipulated the encoding process, the second study (Chapter 3) shifted the focus of the investigation of the production effect as a context-based memory technique to retrieval: Evidence of the use of production as a contextual cue was tested using a contextual change account of directed forgetting. The idea was that instructing participants to remember items based on whether they were studied aloud or silently should reinstate the relevant contextual framework for the aloud items, improving memory of the aloud items and attenuating effects of directed forgetting. The list method procedure from directed forgetting was used, in which people were expected to encode the “aloud” information as a contextual detail.

Two mixed lists were studied, each containing both aloud and silent items, and half of the participants were falsely led to believe that List 1 would not be tested so it could be forgotten. For the aloud items specifically, the contextual change account of directed forgetting would predict that the effect of the forget instruction should be attenuated when participants are instructed to remember the words based on how they studied them (the Aloud versus Silent Differentiation group). This test instruction essentially reinstates the relevant contextual information (for List 1 aloud items in particular) that is hypothesized to support the remembering of the aloud items as a memory aid framework (and which is proposed by the distinctiveness account). In contrast, instructing participants who had been told that they should forget List 1 that they should now remember the words based on which list they had appeared in (the List 1 versus List 2 differentiation group) would not reinstate the relevant contextual information

because list information is not a distinct contextual cue, and therefore this instruction should lead to typical directed forgetting effects.

Specifically, in this list differentiation group, a significant interaction is predicted to occur in memory for items studied aloud: The hit rate for List 1 should be lower in the forget group than in the remember group, whereas the hit rate for List 2 should be higher in the forget group than in the remember group—the typical directed forgetting effects seen in that literature. In contrast, for the Aloud versus Silent Differentiation group, there should be no difference between the forget and remember groups. This predicted contrast between the two groups would support the central hypothesis of this dissertation in that people do establish source memory for studying aloud items and, in this case, when the test is recognition, source memory can then be used as contextual information to overcome the effect of directed forgetting. Like Study 1, production would act as a contextual cue to which all of the aloud items would become associated during study (but selectively without the silent items in this case, since a mixed-list design was used) and reinstating the use of this information would then facilitate retrieval.

The third study (Chapter 4) examined how the distinctiveness of aloud items, as a contextual-based effect, may interfere with the encoding of other items (aloud or silent) when the test is recall. As discussed, previous production effect studies using recall tests have suggested that when people study mixed lists, the distinctiveness—or “uniqueness”—aspect of studying items aloud disrupts list order information during the encoding of silent items, leading to a memory cost for these silent items. Here, the investigation was extended to more specifically examine interference effects caused by studying aloud items—as well as possibly by studying silent items. The goal was to examine the effects of retroactive versus proactive interference that may stem from each type of study. Participants studied (1) four trials each consisting of one

aloud list and one silent list (counterbalanced), (2) two trials each consisting of two aloud lists, and (3) two trials each consisting of two silent lists, in addition to two control trials of only one aloud or one silent list. In the trials where two lists were studied, testing of List 1 or List 2 was alternated across trials such that each list was tested once but the order of testing was random.

In all of the two-list trials, the expectation was that there would be proactive interference from List 1 onto List 2, as well as retroactive interference from List 2 onto List 1, the typical pattern. Relative to studying List 2 silently, however, studying List 2 aloud was expected to invoke greater retroactive interference onto List 1 items, regardless of the study mode of List 1. In contrast, this encoding effect difference was not predicted when examining the possible proactive interference effects of study mode of List 1. The hypothesis is that when a list of aloud items is studied second, the greater distinctiveness of these aloud items would produce a release from buildup of proactive interference from previously studied items because a change of context would occur when List 2 was studied aloud (regardless of the study mode of List 1). This contextual framework for List 2 is then maintained into retrieval when retrieval occurs in close temporal proximity to encoding, improving memory for these aloud items studied later.

To generalize to the case of mixed lists containing both aloud and silent items, this release from proactive interference buildup is hypothesized to occur each time an item is studied aloud, and only silent items would be “released” through the study of later aloud items—the aloud items that are already studied would be maintained because they are all studied within one “list” and associated under a common contextual framework of production. This encoding mechanism for the formation of production as a global context is presumed to be identical in mixed-list designs whether the test is recall or recognition. The retrieval method, however, would influence how the effect of production on memory is revealed, because the studied information is

presented again at retrieval in recognition tests but not in recall tests. These ideas will be discussed in more detail in Chapter 4.

More rationale—including connections to the relevant literature—will be presented in the introductory sections of each study across the next three chapters. Collectively, the three empirical chapters will build the case for the critical role of context in the production effect. The final chapter will then bring this all together, concluding how the production effect can be successfully recast as another instance of the ubiquitous role of context in memory.

Chapter 2 – The Pure-List Production Effect and Statistical Distinctiveness⁴

This chapter will cover Study 1: The examination of whether the use of the distinctiveness heuristic, as a contextual-based memory aid, will boost memory for information studied aloud even in the case of encoding of pure lists. To summarize, the three experiments reported in this study examined the influence of (1) the distinctiveness heuristic in a pure-list paradigm and also (2) statistical distinctiveness during study. As discussed in Chapter 1, the first research question involved the study of designated pure-list aloud or silent items presented in alternation with mixed-list aloud and silent items. Through this manipulation, would the production effect for the pure-list items be enhanced by the co-occurring within-subject manipulation? That is, would participants become aware that the distinctiveness heuristic can be used to aid recognition even for aloud items studied in pure-lists—without being explicitly told so? The second research question addressed whether the encoding of the designated pure-list items in conjunction with mixed-list items would influence the magnitude of the production effect observed, as memory for these pure-list items would effectively change the overall proportion of aloud versus silent items studied. The rationale for both questions is discussed in more detail in the following.

The Distinctiveness Heuristic in Pure-List Designs

The first question pertains to whether, for distinctive processing to be effective, the aloud-silent distinction must be experienced in a mixed-list paradigm, where obvious shifts between the aloud and silent conditions are present such that participants can notice them. Specifically, when studying a pure list of aloud words, would participants' memory for these

⁴ The experiments reported in this chapter have been published in *Memory* (Zhou & MacLeod, 2021). The text has been slightly modified from the published version to fit the structure of this dissertation.

items improve if, either consciously or unconsciously, they implemented the distinctiveness heuristic for the aloud words just as when studying mixed lists? That is, would they come to use memory for whether items had been produced at study to help their recognition, thereby increasing the magnitude of the production effect in a pure-list between-subjects design?

As discussed in Chapter 1, studies have shown that although the production effect is indeed present under between-subjects, pure list designs as well as in within-subject, mixed list designs, it appears to be considerably smaller between subjects (see MacLeod & Bodner, 2017, for a brief review). I hypothesized that this could be due to an inconsistent application of the distinctiveness heuristic in the study of pure lists of aloud items, because the technique would not be obviously useful in these situations. However, on a recognition test, it is entirely possible to use the distinctiveness heuristic to boost memory for these aloud items, because the non-studied distractors can be contrasted to the aloud items due to not having been produced. If the distinctiveness heuristic can be successfully used to improve memory for information studied aloud in pure-list designs up to that of a similar magnitude as in mixed-list designs, there would be support for the central hypothesis that production can act as an associative contextual cue which aids memory whenever information is studied aloud, rather than simply a mechanism that relies on some form of distinction among target items during encoding (such as reading aloud versus silently).

To answer this first question, the aim was to design experiments in which participants are not explicitly instructed that they could use the distinctiveness heuristic to their advantage, since this is not done in a mixed-list paradigm. This was accomplished by inserting a pure-list set of items to be studied either all aloud or all silently, into mixed lists containing both aloud and silent items: One group of participants would study the pure list items all aloud and another

group would study the pure list items all silently, and then be tested on all items. It was then possible to compare whether the production effect for the pure-list items was affected given that participants should appreciate the value of the distinctiveness heuristic from experience with the mixed-list manipulation. The final hurdle was to find a way to differentiate the pure-list items from the mixed-list items; otherwise, participants would still be studying mixed lists but simply with the proportion of aloud and silent items altered from the typical equal split in previous production effect studies.

To solve this problem, participants were given a cover story in which the pure-list items would be presented in a different colour from the mixed-list items. Participants were also told that they did not have to remember the pure-list items because they were irrelevant and would not be tested. It was predicted, however, that participants would have above-chance memory for these items at least if they were studied aloud, since participants could not avoid processing them. In contrast, the pure-list items could be ignored if they were studied silently, and the question was to determine whether participants would still show above-chance memory for these items in such a case. If participants showed above-chance memory for the pure-list items in both groups, then it would be possible to examine the magnitude of the production effect in a between-subjects paradigm where participants should have detected the value of the distinctiveness heuristic because of the co-occurring mixed-list manipulation, thereby answering the original research question.

Therefore, experiments were designed where presentation of the pure-list items alternated with the mixed-list items, and the pure-list items were presented in a different colour. To ensure that such a design per se does not influence the magnitude of the production effect within-subject, given that the predictions rested on having a similar production effect compared to

previous studies, a baseline experiment was conducted first. In Experiment 1, after every aloud or silent item, an irrelevant event was inserted that required no processing: a row of Xs in a colour not used for the to-be-remembered items. In Experiments 2 and 3, the row of Xs was replaced with the pure-list items that participants read either all aloud or all silently. This, then, will be the first time that the within-subject and the between-subjects designs are examined in the same experiment.

Statistical Distinctiveness

This setup also enabled the examination of a second research question: The examination of the influence of relative frequency—or statistical distinctiveness—of the aloud and silent subsets of words at study. Ordinarily, in a mixed list, there are equal numbers of words read aloud and silently, but what happens if this balance is disturbed? There is some limited evidence showing that statistical distinctiveness modulates the size of the production effect. Ozubko and MacLeod (2010) showed that when participants studied one mixed list and one pure list, the production effect for the mixed list was robust for those whose pure list was all silent but was eliminated for those whose pure list was all aloud. Presumably, having an additional pure list of all aloud words reduced the statistical distinctiveness of the aloud words in the mixed list, hence the finding of a reduced production effect. Icht et al. (2014; see also Bodner et al., 2016) directly manipulated the relative frequency of the aloud and silent words in the study list and found a similar pattern of results: The magnitude of the production effect was reduced as the statistical distinctiveness of the aloud words decreased (i.e., as the proportion of aloud words relative to silent words was increased).

As just described, in Experiments 2 and 3, the rows of Xs between successive mixed-list words were replaced with the pure-list words in a different colour; the pure-list words were read

all aloud by one group and all silently by another group, and participants were told that they did not have to remember these items. Would words that were not relevant and that could be forgotten nevertheless influence the magnitude of the production effect with respect to statistical distinctiveness? Specifically, when the irrelevant words were read aloud, would they lower the statistical distinctiveness of the to-be-remembered aloud words in the mixed list (given the prediction that participants would remember them above chance), thereby differentially affecting the production effect relative to when the irrelevant words were read silently, which should increase the statistical distinctiveness? Or would they—simply because they were defined as irrelevant—not influence the magnitude of the production effect under a statistical distinctiveness hypothesis?

In summary, the goals of the present study were twofold. First, the investigation pertained to whether, similar to mixed-list designs, being made aware of the distinctiveness heuristic in a pure-list paradigm—through observing the aloud-silent distinction when pure-list items are inserted into mixed lists—would influence the magnitude of the production effect. Second, the experimental setup allowed the examination of whether changing the relative frequency of the aloud and silent words by the addition of the pure-list words at study would differentially affect the size of the production effect under a statistical distinctiveness account.

Experiment 1

Experiment 1 served as a baseline for the main experiments to follow. Here, each word designated as aloud or silent by its colour (e.g., blue = aloud; white = silent) was followed by a row of Xs in red such that immediate contrasts of aloud and silent items were precluded. No response was required to these red Xs. The goal was simply to prevent the aloud and silent items

from being adjacent to each other, and to determine whether these interleaved red events would disrupt the normal production effect by reducing the aloud/silent contrast on a trial-by-trial basis.

Method

Participants. Participants were 24 undergraduate students (9 men, 13 women; age range: 17-25 years, mean age: 20.6 years, $SD = 2.9$ years; 2 participants declined to provide demographic information) from the University of Waterloo, recruited via the Department of Psychology's research participation system. Ethics approval was obtained from a University of Waterloo Research Ethics Board, and written consent was obtained from all participants (the same applies to Experiments 2 and 3). Participants received either course credit or \$5 in exchange for their participation. Based on the existing studies examining the production effect, this number of participants—frequently used in prior studies—was believed to be sufficient to test for a production effect in the current manipulation, and would allow a comparison of the magnitude of the production effect obtained in this experiment to previous studies.

Apparatus. The experiment was controlled by a PC-compatible computer running a program written in E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA). Study and test trials were presented on a LCD monitor, with responses collected via a standard QWERTY keyboard.

Materials. In both the study and test phases, words were presented in lowercase in the Consolas font, size 36, against a black background. The set of 120 words was the set used in MacDonald and MacLeod (1998); these are listed in Appendix A. Words that formed each condition (i.e. aloud, silent, or new) were selected randomly for each participant. Each participant studied 40 words aloud and 40 words silently, with the sequence of words and conditions randomized. The remaining 40 words were used as distractors on the recognition test,

with the resulting 120 test words randomized anew. This 2:1 ratio of targets to distractors has been commonly used in production effect studies (e.g., MacLeod et al., 2010) but so has the 1:1 ratio (e.g., MacLeod, 2011): A robust production effect is observed in both arrangements.

Procedure. For the study phase, participants were instructed to read the words presented in blue aloud and the words presented in white silently; silent reading was to be done without moving their lips. Each blue or white study word was presented individually for 3 s at the centre of the screen. Between successive aloud or silent words, a row of five red Xs was presented at the centre of the screen for 3 s. Participants were told that no action was required during these red-X displays. A blank period of 500 ms intervened between successive stimuli.

A recognition test immediately followed the study phase. The 80 studied words, intermingled with the 40 distractor words, were presented one at a time in a random order in yellow, so that colour would not serve as a retrieval cue. Using a key press, participants indicated whether they remembered studying each word (the “Y” key for yes; the “N” key for no), taking as long as they needed for each response. Upon response, the word disappeared and, after a 500-ms blank screen, the next word appeared. The entire procedure took under 30 minutes.

Results

The recognition test data are shown in the top row of Table 1 as well as in Figure 1: The dependent measure is hit rate. Primary analyses were conducted using IBM SPSS Statistics (Version 26).⁵ The single false alarm rate was low, indicating overall good memory for the studied words. As expected, a paired-samples *t*-test revealed that memory was significantly

⁵ Estimates of Hedges’ *g* were computed according to the formulae provided in Cumming (2012). For confidence intervals of the effect sizes, the boundaries for the noncentrality parameter were computed first using the MBESS package in R Version 3.6.3 (Kelley, 2007); then (except for paired *t*-tests) the corresponding confidence intervals for the effect sizes were computed according to the formulae provided in Smithson (2003), Chapters 4 and 5. For paired comparisons, confidence intervals for the effect sizes were computed according to Algina and Keselman (2003).

better for the aloud words than for the silent words, $t(23) = 5.77, p < .001, g_{av} = 1.01, 95\% \text{ CI } [0.575, 1.51]^6$, a quite typical production effect of .178. Clearly, then, interleaving the repeated red stimuli among the aloud and silent words at study without requiring any action from participants did not reduce the normally observed production effect. This set the stage for changing the interleaved items from easily-ignored Xs to the pure-list words, thereby allowing the investigation of both research questions.

Table 1

Proportions of “Yes” responses, corresponding to hits for studied targets and false alarms for unstudied distractors in Experiments 1-3.

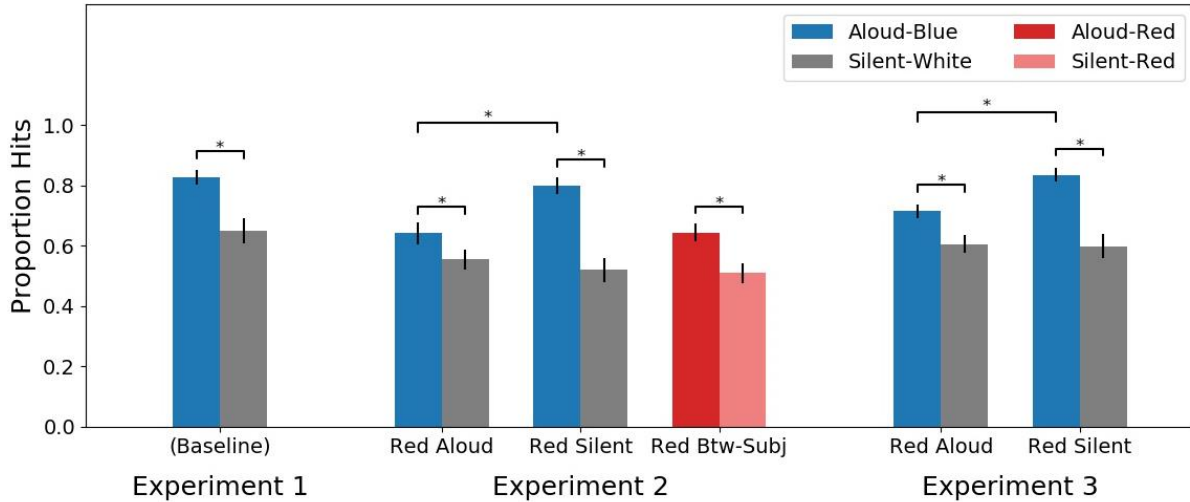
Experiment	Aloud	Silent	Distractor	Red	Red Distractor
1 – red Xs	.827 (.024)	.649 (.043)	.197 (.034)		
2 – red aloud	.642 (.037)	.554 (.034)	.218 (.031)	.644 (.030)	.157 (.020)
2 – red silent	.799 (.029)	.520 (.042)	.233 (.041)	.510 (.033)	.229 (.026)
3 – red aloud	.715 (.022)	.605 (.030)	.151 (.025)		
3 – red silent	.835 (.023)	.599 (.040)	.189 (.031)		

Note. Standard errors shown in parentheses below each mean.

⁶ Effect size confidence intervals for t -tests were calculated based on d (see Cumming, 2012, chapter 11).

Figure 1

Proportions of hits for studied targets in Experiments 1-3.



Note. Error bars represent standard error. Significant differences are indicated with an asterisk.

Experiment 2

Experiment 1 demonstrated that the production effect remained robust despite the interleaved irrelevant Xs. Thus, this simple manipulation interrupting the usual switching back and forth between aloud and silent items in a mixed list did not alter the effect. In Experiment 2, the red Xs were replaced with the pure-list words in red. One group read all of these red words aloud whereas the other group read all of them silently.

If the red words are simply compartmentalized as disruptor items, then no change would be expected to occur in terms of the magnitude of the mixed-list production effect. If, however, participants are driven by this manipulation to actually process and remember the red words, this would create a situation where it would be possible to examine the between-subjects production effect for the red words, as well as to examine the two different situations in which there now would be more aloud than silent words when the irrelevant words were read aloud, or more silent than aloud words when the irrelevant words were read silently. In the first case, the aloud words

would be made less statistically distinctive, and consequently the production effect for the to-be-remembered blue and white words would be reduced. In the second case, the aloud words would be made more statistically distinctive, and the production effect for the to-be-remembered blue and white words would increase.

Given that the aim was to test memory for both the mixed-list and pure-list words, there was no ideal order for examining both. Therefore, first in Experiment 2, memory for the pure-list red words was examined before testing memory for the mixed-list words. This order ensured that testing of the mixed-list words did not interfere with the pure-list words while also not hindering any possible effect of the distinctiveness heuristic at test, since the hypothesis was that participants would become aware of the strategy through the study phase alone (i.e., when remembering the aloud pure-list items, they would use the heuristic to distinguish between the aloud and distractor items). In Experiment 3, which partially served as a replication of Experiment 2, participants were correctly informed that they would not be tested on the pure-list red words, so that it was possible to confirm the results from testing only the mixed-list words.

Method

Participants. The participants were 48 undergraduate students (13 men, 34 women; age range: 17-33 years, mean age: 20.3 years, $SD = 2.6$ years; 3 participants declined to provide their age, 1 participant declined to provide demographic information) from the same source as Experiment 1. Participants received either course credit or \$5 in exchange for their participation. Half of these participants were randomly assigned to the red-aloud condition and half to the red-silent condition. (Although an explicit power analysis was not performed for this experiment, the results of Experiment 2 will be used to perform power analyses for Experiment 3, which partially serves as a replication of Experiment 2.)

Apparatus and stimuli. The apparatus was the same as in Experiment 1. The set of words used in Experiment 1 was augmented for Experiment 2; the additional words are also shown in Appendix A. The study and test lists were constructed as in Experiment 1, but with the addition of 79 red words inserted between successive aloud and silent words during study. An additional 40 words were included in the stimulus set to accommodate the test of the pure-list red words. The entire set of words was randomized anew for each participant.

Procedure. During the study phase, there were blue and white words to be treated as in Experiment 1. Instead of red Xs, however, here there were red words presented between successive blue and white words, creating the red-aloud and red-silent pure list conditions. In the red-aloud condition, participants were told to read all of the red words aloud; in the red-silent condition, participants were told to read all of the red words silently. Participants were explicitly told that only the words presented in blue and white would be tested. In fact, though, an incidental recognition test for the red words occurred immediately after the study phase and prior to the recognition test of the intentionally learned blue aloud and white silent words. The 79 red words and the additional 40 distractor words were presented in a random order in yellow during the test. The test was conducted in the same manner as the blue-white words test. Following this was the test for the blue and white words, which was conducted exactly as in Experiment 1.

Results

The results are shown in rows 2 and 3 of Table 1 as well as in Figure 1.

Test of the red words

False alarm rates for the red words were low, indicating that participants had very good memory for the red words despite being told that they would not be tested. The false alarm rate in the red-silent condition, however, was significantly higher than that in the red-aloud condition,

$t(46) = 2.21, p = .032, g = 0.628, 95\% \text{ CI } [.053, 1.21]$. The key result was the hit rate for the red words. Recognition of them was significantly better when they were read aloud (.644) as opposed to silently (.510), $t(46) = 2.99, p = .004, g = 0.850, 95\% \text{ CI } [.267, 1.45]$, indicating a large between-subjects production effect of .134, occurring in the same experiment as the within-subject production effect to be reported next. It is noteworthy that the magnitude of this effect was similar to that of the effect usually seen under the mixed-list design and considerably larger than the effect usually seen under the pure-list design.

Test of the blue and white words

False alarm rates were low for both groups—red-aloud and red-silent—and did not differ, $t(46) = 0.303, p = .763, g = 0.083, 95\% \text{ CI } [-0.479, 0.652]$, evidence of good memory for the studied words. A 2 (blue aloud/white silent) \times 2 (red aloud/red silent) mixed ANOVA was performed to examine how the magnitude of the production effect was influenced by whether participants read the red words aloud or silently. Recognition of the blue (aloud) words (.720) was better than recognition of the white (silent) words (.537), $F(1, 46) = 59.75, MSE = 0.014, p < .001, \eta_p^2 = 0.565, 90\% \text{ CI } [0.394, 0.667]$ ⁷, a quite typical within-subject production effect of .183, and very similar to that of Experiment 1. The main effect of the between-subjects condition (red aloud/red silent) was not significant, $F(1, 46) = 1.89, MSE = 0.048, p = .176, \eta_p^2 = 0.039, 90\% \text{ CI } [0, 0.160]$.

Critically, though, the interaction was significant, $F(1, 46) = 16.33, MSE = 0.014, p < .001, \eta_p^2 = 0.262, 90\% \text{ CI } [0.094, 0.412]$, indicating that the magnitude of the production effect differed according to whether the red words had been read aloud or silently. Given the significant

⁷ For effect sizes represented by the partial eta squared statistic, 90% confidence intervals were computed instead of 95% confidence intervals since partial eta squared cannot be negative (see Steiger, 2004).

interaction, the simple main effects were examined next. Paired-samples t -tests with Bonferroni correction showed a significant production effect both in the red-aloud condition, $t(23) = 2.48$, $p = .021$, $g_{av} = 0.489$, 95% CI [0.076, 0.920], and in the red-silent condition, $t(23) = 8.80$, $p < .001$, $g_{av} = 1.53$, 95% CI [1.00, 2.15]. The effect was, however, significantly larger in the red-silent condition (.279) than in the red-aloud condition (.088), $t(46) = 4.04$, $p < .001$, $g = 1.15$, 95% CI [0.547, 1.77]. As the data in Table 1 show, the production effect was significantly larger in the red-silent condition than in the red-aloud condition because memory for the blue aloud words was significantly higher when the red words had been read silently, $t(46) = 3.33$, $p < .001$, $g = 0.943$, 95% CI [0.358, 1.56]. There was, however, no significant difference in memory for the white silent words as a function of whether the red words were read aloud or silently, $t(46) = 0.639$, $p = .526$, $g = 0.180$, 95% CI [-0.384, 0.750].

Comparing Experiments 1 and 2

Because Experiments 1 and 2 were identical in every respect apart from the critical change from red Xs to red words, performance in the two experiments were compared. First, the analyses for the to-be-remembered blue aloud items. A one-way ANOVA (with Bonferroni correction) with three levels—red Xs (from Experiment 1), red-aloud, and red-silent (both from Experiment 2)—revealed a significant effect, $F(2, 69) = 10.62$, $MSE = 0.022$, $p < .001$, $\eta_p^2 = 0.235$, 90% CI [0.091, 0.353]. Specifically, memory for the blue aloud words was poorer in the red-aloud condition than in the baseline red Xs condition, $t(46) = 4.19$, $p < .001$, $g = 1.18$, 95% CI [0.586, 1.819]. In contrast, there was no difference in memory for the blue aloud words when comparing the red-silent condition with the red Xs baseline, $t(46) = 0.740$, $p = .463$, $g = 0.209$, 95% CI [-0.355, 0.779]. Next, performance on the to-be-remembered white silent words was similarly examined; there was no significant difference, $F(2, 69) = 2.85$, $MSE = 0.038$, $p = .064$,

$\eta_p^2 = 0.076$, 90% CI [0, 0.174]. These results confirm that the decreased magnitude of the production effect in the red-aloud condition was due to a decrease in memory for only the blue aloud words. Reading the red words silently did not result in a significant change relative to the red Xs baseline.

In summary, in Experiment 2, a quite large between-subjects production effect was observed for the red words relative to what has typically been observed in a pure-list experimental design. This finding supports the prediction that inserting a pure-list paradigm within a mixed-list paradigm allowed participants to observe the aloud-silent distinction and to use the distinctiveness heuristic even in a pure-list design, consequently increasing the magnitude of the pure list production effect compared to previous studies. In addition, Experiment 2 demonstrated that inserting the pure-list red words between successive blue aloud and white silent words differentially affected the magnitude of the within-subject production effect: Relative to the Experiment 1 baseline, the production effect decreased when the red words were read aloud but was not significantly altered when the red words were read silently. This is partially consistent with the hypothesis regarding the influence of statistical distinctiveness on the size of the production effect.

Our data fit the pattern of data reported by Icht et al. (2014) and by Bodner et al. (2016): When the red words are read aloud, there is effectively an overall greater number of aloud words than silent words, thereby making the aloud words less distinctive and diminishing the production effect. It is surprising, then, that when the red words were read silently the production effect was not enhanced despite there being as a result relatively fewer aloud words overall, which would be expected to make the aloud words more distinctive. Experiment 3 sought to replicate the latter findings.

Experiment 3

In Experiment 2, to address the first question with respect to pure-list experimental designs, memory for the red words from the pure lists was tested prior to testing the words from the mixed lists. It is possible that this could have interfered with retrieval of the words from the mixed lists in some way, which would affect the interpretation of the results in terms of representing statistical distinctiveness. To address this issue, in Experiment 3 only memory for the mixed-list words were tested. Experiment 3 should therefore provide a straightforward replication of the findings of Experiment 2 regarding statistical distinctiveness.

Method

Participants. The participants were 48 undergraduate students (12 men, 35 women; age range: 17-23 years, mean age: 19.1 years, SD = 1.3 years; 1 participant declined to provide demographic information) from the University of Waterloo, recruited as previously described. Half of these participants were randomly assigned to the red-aloud condition and half to the red-silent condition. A power analysis was performed using G*Power Version 3.1.9.4 (Faul et al., 2007) to determine a suitable sample size based on the results of Experiment 2. Based on $\alpha = .05$ and power = 0.95, and using the effect sizes from the mixed ANOVA and one-way ANOVA for the aloud items⁸ in Experiment 2, the appropriate sample size needed to satisfy these parameters was estimated to be $N = 18$ per group. Therefore, the sample size of $N = 24$ per group is adequate for investigating the main question in the current experiment.

Apparatus, stimuli, and procedure. The apparatus, stimuli, and procedure were the same as in Experiment 2 except that participants were only tested on the mixed-list words. To maintain consistency with Experiment 2, participants were now correctly told at the start of the

⁸ Only the aloud items were examined here because in referencing the results from Icht et al., (2014), differences across groups for memory of the silent items were not expected.

experiment that the words in red would not be tested. Forty words were removed from the set of words used in Experiment 2 as a result (see Appendix A). The set of words remaining was randomized anew for each participant.

Results

The recognition test data are shown in the bottom two rows of Table 1 as well as in Figure 1. False alarm rates were low in both groups (red-aloud and red-silent) and did not differ, $t(46) = 0.946, p = .349, g = 0.269, 95\% \text{ CI } [-0.297, 0.840]$, evidence of good memory for the studied words. As in Experiment 2, a 2 (blue aloud/white silent; within) \times 2 (red aloud/red silent; between) mixed ANOVA was performed first to examine the magnitude of the production effect as a function of whether participants read the red words aloud or silently. Recognition of the blue (aloud) words (.775) was better than recognition of the white (silent) words (.602), $F(1, 46) = 78.85, MSE = 0.009, p < .001, \eta_p^2 = 0.632, 90\% \text{ CI } [0.476, 0.719]$, a quite typical within-subject production effect of .173, and a pattern very similar to that of Experiments 1 and 2. Similar to Experiment 2, the main effect of the between-subjects condition (red aloud/red silent) was not significant, $F(1, 46) = 2.39, MSE = 0.033, p = .129, \eta_p^2 = 0.049, 90\% \text{ CI } [0, 0.176]$.

As in Experiment 2, the interaction was significant, $F(1, 46) = 10.65, MSE = 0.033, p = .002, \eta_p^2 = 0.188, 90\% \text{ CI } [0.045, 0.341]$, again indicating that the magnitude of the production effect for the blue and white words differed according to whether the red words were read aloud or silently. The relevant simple main effects were therefore probed. Paired-samples t -tests with Bonferroni correction showed a production effect for both conditions, red-aloud: $t(23) = 4.85, p < .001, g_{av} = 0.828, 95\% \text{ CI } [0.426, 1.27]$; and red-silent: $t(23) = 7.45, p < .001, g_{av} = 1.42, 95\% \text{ CI } [0.892, 2.04]$. The effect was, however, significantly larger in the red-silent condition (.236) than in the red-aloud condition (.110), $t(46) = 3.26, p = .002, g = 0.923, 95\% \text{ CI } [0.341, 1.54]$.

Again just as in Experiment 2, this difference was entirely due to memory for the blue aloud words being significantly better when the red words were read silently than when they were read aloud, $t(46) = 3.77, p < .001, g = 1.07, 95\% \text{ CI } [0.476, 1.69]$. There was no significant difference in memory for the white silent words as a function of whether the red words were read aloud or silently, $t(46) = 0.125, p = .901, g = 0.034, 95\% \text{ CI } [-0.528, 0.600]$.

Comparing Experiments 1 and 3

We begin with analyses for the blue aloud items. A one-way ANOVA (with Bonferroni correction) with three levels—red Xs (from Experiment 1), red-aloud, and red-silent (both from Experiment 3)—revealed a significant effect, $F(2, 69) = 8.47, MSE = 0.013, p < .001, \eta_p^2 = 0.197, 90\% \text{ CI } [0.062, 0.314]$. Specifically, this difference resulted from poorer memory for the blue aloud words in the red-aloud condition relative to the red Xs baseline condition, $t(46) = 3.42, p < .001, g = 0.978, 95\% \text{ CI } [0.384, 1.58]$, whereas there was no difference in memory for the blue aloud words when comparing the red-silent condition with the red Xs baseline, $t(46) = 0.249, p = .805, g = 0.068, 95\% \text{ CI } [-0.494, 0.638]$. In contrast to the effect on the blue aloud words, the difference in memory for the white silent words between Experiments 1 and 3 was not significant, $F(2, 69) = 0.52, MSE = 0.034, p = .598, \eta_p^2 = 0.015, 90\% \text{ CI } [0, 0.069]$.

Comparing Experiments 2 and 3

Experiments 2 and 3 differed only in whether the red words were tested. Consequently, it makes sense to compare their results for the blue and white words directly. Toward this end, a 2 (blue aloud/white silent; within) \times 2 (red-aloud/red-silent; between) \times 2 (Experiment 2/3; between) mixed ANOVA (with Bonferroni correction) was carried out. The three-way interaction was not significant, $F(1, 92) = 1.11, MSE = 0.011, p = .295, \eta_p^2 = 0.012, 90\% \text{ CI } [0,$

0.072], demonstrating the consistency of the results in Experiments 2 and 3.⁹ As confirmation, this analysis also revealed an overall red-aloud/red-silent by blue aloud/white silent interaction, $F(1, 92) = 26.97$, $MSE = 0.011$, $p < .001$, $\eta_p^2 = 0.227$, 90% CI [0.111, 0.339], reaffirming that the between-subjects manipulation of red-aloud versus red-silent influenced the size of the within-subject production effect in the collapsed data. There were no other significant two-way interactions or main effects.

Both Experiments 2 and 3 showed that changes in the magnitude of the production effect between the red Xs baseline and the experimental conditions were due to changes in memory for only the blue aloud words. Specifically, the changes in the size of the production effect were due to participants having poorer memory for the blue aloud words in the red-aloud condition, which was evident in both experiments. This difference is attributed to the increased relative frequency of aloud words in the red-aloud condition: Participants were remembering the red words even though this was not required. Although participants also showed some memory for the red words in the red-silent condition, which should have correspondingly changed the relative frequency of the silent words in these conditions, as in Experiment 2 no significant influence on the magnitude of the production effect for the white silent words was found.

In summary, the results for memory of the blue aloud and white silent words in Experiment 3 replicated those of Experiment 2, producing consistent within-subject production

⁹ A supporting Bayesian repeated measures ANOVA, conducted in JASP Version 0.14 (JASP Team, 2020) using uniform priors, revealed anecdotal evidence for the null hypothesis (Lee & Wagenmakers, 2013), $BF_{01}=2.273$. This may be a consequence of the blue aloud and white silent items having been tested second in Experiment 2, yielding somewhat lower performance numerically compared to Experiment 3. Because there is agreement between the individual analyses in Experiments 2 and 3, this was seen as a reasonable basis to conclude that the results of Experiments 2 and 3 were similar. (Note that the error percentage from this analysis was 6.886%, indicating that the Bayes Factor will change slightly when the analysis is repeated; however, this is within the acceptable error percentage limit: see van Doorn et al., 2020.)

effects. That the way in which the red words were processed also influenced the magnitude of the production effect for the blue versus white words—the effect was smaller when the red words were read aloud—fits nicely with the statistical distinctiveness explanation: Reading the red words aloud essentially makes the blue aloud words less distinctive. In the following, further elaboration is provided for why the magnitude of the production effect may not have increased in the red-silent condition, as well as why the between-subjects production effect observed in Experiment 2 appears to be larger than what has typically been observed.

Discussion

Three experiments integrated the examination of two dimensions of distinctiveness that were hypothesized to influence the magnitude of the production effect. First, the distinctiveness heuristic was extended to the pure-list experimental design, to test whether the magnitude of the production effect in a between-subjects setup would increase when participants are shown that they can use the aloud-silent distinction to their advantage when remembering the study items under this type of design. Here, a situation was also created where both within-subject and between-subjects production effects could be examined in the same experiment. Second, the effect of statistical distinctiveness was examined using situations in which the number of aloud and silent items at study were unequal: This design is atypical of past production effect studies. To accomplish these goals, in the present experiments additional words that participants read either all aloud or all silently were inserted between successive to-be-remembered aloud and silent words, and these inserted pure-list items were distinguished from the mixed-list items by providing participants with a cover story that these additional words would not be tested. The results were compared between these experimental groups as well as with a baseline group not required to perform any action when additional stimuli were presented.

The results support the prediction made for the first research question—that the magnitude of the production effect would increase relative to what has been shown in previous production effect studies when participants realize, consciously or unconsciously, that they can use the distinctiveness heuristic even in a pure-list experimental design. This hypothesis was able to be tested in the current setup given that participants showed above-chance memory for the additional red words both when they were read aloud and when they were read silently. The present study therefore allowed, for the first time, the investigation of the between-subjects and within-subject production effect magnitudes in the same experiment. Aloud versus silent was manipulated within-subject using the blue versus white cues; it was manipulated between-subjects by having one group read all of the red words aloud and the other group read all of the red words silently.

In the present data, the production effect was numerically somewhat smaller between-subjects than within-subject (.134 between-subjects, and .183 in Experiment 2/.173 in Experiment 3 within-subject). However, the between-subjects effect here, with an effect size of 0.864, is considerably larger than what has been reported in previous between-subjects studies: In his meta-analysis, Fawcett (2013) reported an average between-subjects effect size of 0.37 with a 95% confidence interval of [0.16, 0.57] across twelve experiments that employed three different testing methods. As explained, demonstrating the aloud-silent distinction in the present experimental design, especially in the group reading the red words aloud, was suspected to have invoked greater use of the distinctiveness heuristic at test (Dodson & Schacter, 2001). This would mean that when participants observe distinctiveness at encoding they become more likely to use the distinctiveness heuristic. They do not need to be explicitly instructed to use this heuristic—and, again, to be clear, the claim is that they do not even need to be consciously aware

of using this heuristic—to show a robust production effect even in the pure-list paradigm. The current findings thus provide support for the hypothesis that previous studies may have shown smaller production effects in pure-list relative to mixed-list designs due to the inconsistent use of the distinctiveness technique in pure-list designs, because it is not obviously useful in these situations.

With regard to the central question of this dissertation, the hypothesis was that production would act as a contextual framework specifying the encoding method to which the pure-list aloud items would become associated, which can then be used to aid the retrieval of these aloud items; the non-studied distractor items on the recognition test would be differentiated by virtue of not having been produced and hence do not constitute part of this framework. This is, in essence, an extension of the utilization of the distinctiveness heuristic within mixed-list designs. The present findings offer support for this hypothesis: Through the concurrent encoding of mixed-list items, participants appear to have discovered that production can act as a contextual cue whenever there is information that is studied aloud, and invoking the use of this contextual cue then facilitates later retrieval of the aloud information. The distinctiveness heuristic thus presents itself as a standalone context-based memory technique for remembering produced information, without relying on the presence of different encoding processes (such as reading silently) to elicit the desired effect.

Second, the pattern of data partially supports the statistical distinctiveness account presented by Icht et al. (2014; see also Bodner et al., 2016), where it was shown that the size of the production effect was reduced as the statistical distinctiveness of the aloud words decreased (i.e., increasing the number of aloud words relative to silent words at study), and vice versa. In the present experiments, a similar pattern was found in the groups of participants who read the

additional red words aloud. These participants had good memory for the red words despite being told that those words would not be tested. Effectively, reading the red words aloud, even though participants were told that they were irrelevant and would not be tested, increased the proportion of aloud words compared to silent words.

Further supporting evidence in Experiment 2 showed that the mean hit rate for the red words when they were read aloud was almost the same as the mean hit rate for the to-be-remembered blue aloud words, indicating that participants actually processed and remembered these red words effectively. Because participants in the red-aloud condition were actually remembering a greater number of aloud words compared to silent words, the aloud words were less statistically distinct, so the observed decrease in memory for the aloud words would be expected according to this extension of the distinctiveness account. The main finding was replicated in Experiment 3: The magnitude of the production effect was reduced in the red-aloud condition because memory for the to-be-remembered blue aloud items declined relative to the baseline at the same time as there was no corresponding change in memory for the silent words.

The findings do not, however, support the statistical distinctiveness account in the opposite direction: The account calls for a greater production effect when the aloud words are more statistically distinct than the silent words; that is, when there are fewer aloud words than silent words at study. In the red-silent conditions, participants also showed good memory for the red words. Because participants were now effectively studying more silent words than aloud words, under a statistical distinctiveness framework an increase in the size of the production effect would have been expected, where memory for the aloud words should have increased relative to the baseline. Instead, no difference in performance was found between the red-silent conditions relative to the baseline.

This difference between the red-silent and the red-aloud conditions cannot be attributed to more attention having been devoted to the red-aloud items such that they became associated as one group with the mixed-list aloud items, whereas the red-silent items were treated as a “to-be-ignored” third group and thus did not affect statistical distinctiveness. In the red-silent condition, participants showed memory for the red items comparable to that for the white-silent items (see MacDonald & MacLeod, 1998, for a related result in Experiment 3 of their study: Where participants could truly ignore the non-spoken items, they showed no recognition memory for these items). In addition, although the numerical differences in the magnitudes of the production effect across the three experiments showed a trend in the direction of a greater production effect in the red-silent conditions compared to the baseline (.279 and .236 versus .178), there is one main difference between the present results and the results of Icht et al. (2014): They showed a memory boost for the aloud words in the condition with fewer aloud items (aloud 20%) relative to their other conditions, whereas the numerical difference in the present results derived mainly from reduced memory for the silent items. Moreover, this difference was not significant across experiments and Icht et al. (2014), using a recognition test, also found no difference in memory for silent items.¹⁰ It therefore appears that the change predicted by the statistical distinctiveness account may not be as robust in the case where there are more silent words relative to aloud words, a possibility worthy of further study in the future.

Because the current data only provide partial evidence to support the influence of statistical distinctiveness on the production effect, I also offer an alternative context-based explanation for these findings (albeit post-hoc) as context may in fact be a more suitable

¹⁰ A supporting Bayesian one-way ANOVA examining just the white silent items across all three experiments, conducted using uniform priors in JASP, revealed a Bayes factor indicating moderate evidence for the null hypothesis (Lee & Wagenmakers, 2013), $BF_{01}=3.069$ (with error percentage < 0.001%).

explanation for the present findings. This hypothesis concerns the consistency of the current results with the predictions of the cue overload hypothesis—one that has been presented within the existing literature examining the effects of context on memory (e.g., Earhard, 1967; Rutherford, 2004; Smith & Manzano, 2010; Watkins & Watkins 1975). The cue overload hypothesis states that as the number of items that become associated with a particular context increases, the specificity of the contextual association is reduced, thus increasing the search set and decreasing the effectiveness of contextual cueing at retrieval. For example, in Smith and Manzano (2010), participants encoded 30 words presented against 2, 10, or 30 video clips played in the background, with one clip played per word studied; participants were then asked to recall the words while the video clips were played again during retrieval. The video clips thus acted as contextual cues that varied in their degree of specificity for aiding memory of the associated words. Memory was highest in the condition where the most video clips (30) were used and lowest in the condition with the least video clips (2). This pattern of findings was attributed to an effect of gradual cue overloading, occurring as each video clip was associated with a increasing number of words: In the 30 video clips condition, each video clip was uniquely associated with one word, while in the 2 video clips condition, each video clip had to be associated with 15 words. As the number of associations increased, the use of the video clips as effective memory cues became more difficult due to the decreasing specificity of the associations. Recently, it has been suggested that the well known von Restorff effect (von Restorff, 1933; see Hunt, 1995; Wallace, 1965, for reviews)—essentially, enhanced memory for uniquely isolated items—also falls under the umbrella of cue overloading effects: The unique item can be understood as possessing a distinct context while any remaining items are all associated with one shared context (Chee & Goh, 2018).

The current experiments showed that increasing the *absolute* number of aloud items encoded relative to the baseline (i.e., when the pure-list items were studied aloud) reduced the magnitude of the production effect by decreasing overall memory for the aloud items. In contrast, decreasing the *relative* number of aloud items encoded (i.e., when the pure-list items were studied silently)—while the actual number of aloud items remained the same as the baseline—did not affect the magnitude of the production effect: Memory of both the aloud and silent items remained unchanged. As discussed, these findings are inconsistent with the statistical distinctiveness hypothesis presented in Icht et al. (2014), who predicted that the magnitude of the production effect should vary according to changes in the *relative* frequency of the aloud versus silent items within a study set. Instead, the apparent partial influence of statistical distinctiveness when the pure-list items were studied aloud can be plausibly attributed to an overloading of the production contextual framework when more aloud items were studied, as the memory cue provided by production became less unique. Conversely, when more silent items were studied, the number of aloud items remained the same as the baseline, and no overloading effect would be expected. Production thus possesses the cue overloading characteristic associated with context-based memory effects, which offers further support for the central hypothesis that the production effect belongs to this domain of memory effects.

One important note for the present experiments, however, is that when the additional pure-list items were studied silently, although there was no change in the number of aloud items studied, there were more items encoded in total than was true for the baseline. To my knowledge, there are currently no existing studies that have examined the cue overloading effect in this type of situation: The total number of studied items has always remained constant (this was also the case in Icht et al., 2014). Effects of cue overloading due to changes in only the total number of

studied items—but not the number of items associated with a particular contextual cue—remains a key question requiring future examination.

An interesting connection to the directed forgetting literature (see MacLeod, 1998, for a review) is that, although participants were instructed that they did not need to remember the irrelevant red words, memory for the red words was similar to that for the to-be-remembered blue aloud and white silent words respectively in the red-aloud and red-silent conditions in Experiment 2. The method of the current study resembled the item-method directed forgetting paradigm where a directed forgetting effect has been consistently found when using recognition tests. Accordingly, worse memory would have been expected for the red words compared to the to-be-remembered words in Experiment 2, even if participants were able to remember them above chance level. Why was this not the case? Indicating at the time of test whether an item had been a “remember” or “forget” item at study should have no effect on the magnitude of directed forgetting (Taylor et al., 2018), so it appears that the absence of a directed forgetting effect was due to the nature of the task affecting the encoding phase.

Selective rehearsal of the to-be-remembered items, but not of the to-be-forgotten items, is the most supported explanation of the item-method directed forgetting effect (see MacLeod, 1998; Tan et al., 2020). When the irrelevant red words were studied aloud, it would have been impossible not to rehearse these words at all, although it may be expected that the red words were still rehearsed less than the blue aloud words. Compounding the mystery is the fact that no directed forgetting was seen when the red words were studied silently despite it being possible not to rehearse these words—or perhaps even not to study them at all. Moreover, the red words would have been expected to add to the memory load, reducing overall performance relative to Experiment 1: In directed forgetting, the presence of to-be-forgotten words in a list has been

shown to reduce memory for the to-be-remembered words relative to a list containing only to-be-remembered words (e.g., Muther, 1965). However, memory for the to-be-remembered white silent words in Experiments 2 and 3 was similar to the Experiment 1 baseline. These curious violations of the selective rehearsal account warrant further examination in the future.

One might argue that the greater-than-typical production effect for the pure-list items reflects not usage of the distinctiveness heuristic but rather an effect of incidental versus intentional learning on the silent items. In the generation effect literature it has been suggested that, compared to intentional learning, incidental learning may result in a greater generation effect (e.g., Watkins & Sechler, 1988). The idea is that encoding of the non-generated items is better under intentional than incidental learning whereas memory for the generated items remains similar in the two procedures. In the present experiment, consequently, it is possible that participants showed poorer memory for the silent pure-list items than they would have had those items been intentionally encoded, while there was no effect of distinctiveness on the aloud items.

Unfortunately, there currently are no studies directly comparing intentional versus incidental learning in the production effect. But in an experiment embedding the production effect within the item-method directed forgetting paradigm, Hourihan and MacLeod (2008) showed a greater production effect for the items that participants were told to forget relative to those that they were instructed to remember, as a result of reduced memory for the silent “forget” items. However, the two main differences between the Hourihan and MacLeod (2008) study and the present study are that (1) they used an entirely mixed-list paradigm whereas the present study also aimed to examine pure-list effects, and (2) directed forgetting was not observed in Experiment 2. Further work is needed to specifically investigate the intentional versus incidental learning hypothesis.

The experimental design was set up to distinguish the pure-list items from the mixed-list items. Given the similarity in the pure-list and mixed-list data in Experiment 2, however, it cannot be ruled out that participants may have integrated the pure-list and mixed-list items into a longer mixed list, rather than processing the pure lists as separate entities. The belief was that initially testing memory for the pure-list items in Experiment 2 immediately reminded participants of their status as a separate group of words. Future studies should continue to devise alternative ways of introducing the distinctiveness heuristic to participants without doing so explicitly.

There is one further consideration. In Experiment 1, inserting the same design of a row of red Xs between successive to-be-studied items may have resulted in participants growing accustomed to this stimulus, making it easier to ignore and thus providing participants some extra rehearsal time compared to Experiments 2 and 3. A future study could try using differing numbers of red Xs (and perhaps asking participants to indicate the number of Xs shown) to mitigate this potential issue.¹¹

In summary, these experiments demonstrate that the pure list, between-subjects production effect can be enhanced if the distinctiveness of the pure aloud items is made more apparent by a concurrent mixed list, within-subject manipulation. The finding that production serves as a useful memory tool even in pure lists supports the idea that production acts as a context-based memory tool, and association of produced items—encoded in any circumstance—with this framework facilitates later retrieval of these produced items. Although these experiments may also provide evidence partially in support of the statistical distinctiveness account, an alternative context-based explanation was also discussed, positing that even the

¹¹ The authors thank Reviewer 1 (from the peer-review process for publication) for suggesting this idea.

addition of words that participants do not have to remember influences the number of aloud and silent words studied, thereby altering the magnitude of the production effect through cue overloading. Taken together, these findings offer support for a context-based distinctiveness explanation of the production effect.

Chapter 3 – Examining the Production Effect using Directed Forgetting

This chapter will cover Study 2—using a directed forgetting procedure to investigate the distinctiveness of produced information as a contextual-based effect in recognition memory. To summarize, Study 2 aims to reinforce that within recognition memory, production does indeed function as a contextual framework, in that reinstating the use of this information enhances memory only of aloud items—and not of silent items—intentionally forgotten. Specifically, the prediction is that the distinctive processing afforded by the produced items represents a global contextual framework that becomes associated with (only) these items during encoding; if these produced items were initially forgotten but subsequently need to be retrieved, this framework would serve to facilitate their retrieval—if one is prompted to use this contextual cue.

The design of the present experiment was informed by the contextual change hypothesis originally proposed by Sahakyan and Kelley (2002) as an account of list-method directed forgetting but also applied more widely to other phenomena (e.g., retrieval-induced forgetting, Jonker et al., 2013, 2015). Their procedure used context reinstatement after intentional forgetting, precisely the approach that is required to address the research question in the present study.

Contextual Change in Directed Forgetting

The contextual change hypothesis uses a framework of global (i.e., list-wide) set differentiation due to intentional forgetting (Bjork, 1970) to explain the effects of directed forgetting in the list-method procedure. In this procedure, participants typically are given two lists of items to study. After studying List 1 but prior to studying List 2, some participants are falsely led to believe that the items from List 1 will not be tested; in fact, all participants are tested on both lists. Relative to a control (“Remember”) group not instructed to forget List 1,

participants in this “Forget” group show worse memory for the List 1 items coupled with enhanced memory for the List 2 items (see MacLeod, 1998; Sahakyan et al., 2013, for reviews). The contextual change hypothesis proposes that when participants are told that they do not need to remember List 1, a mental context change occurs such that the new information from List 2 becomes associated with a set of contextual features different from those associated with List 1, the result of sampling of new contextual cues during the study of List 2. DuBrow et al. (2017) present a review of evidence suggesting that mental context shifts can occur quickly due to processes that include goal-oriented behaviour, which is certainly the case with intentional forgetting.

When participants are given a retrieval test with little or no temporal delay after study, retrieval occurs within the same mental context as the study phase of List 2, facilitating memory for List 2 because the same contextual cues remain in place. In contrast, the relevant contextual cues from List 1 are less available, leading to a contextual mismatch between List 1 and retrieval and consequently to forgetting of the List 1 information.¹² If, however, participants are prompted to try to remember and use the relevant contextual information from the study of List 1 at retrieval, then the effect of directed forgetting should be attenuated to at least some extent. The idea is that participants are capable of remembering the relevant contextual cues associated with the study of List 1 that would facilitate memory for List 1 items (i.e., recollection-based cues) but that they tend not to do so without explicit instructions (Smith, 1979).

¹² Under the contextual change hypothesis, then, list-method directed forgetting can also be characterized as an effect of release from proactive interference buildup (Bjork, 1970; Bjork & Bjork, 1996; Kliegl & Bauml, 2021)—a forget instruction after studying List 1 induces a mental context change prior to studying List 2, removing effects due to proactive interference from List 1 onto List 2 and thereby enhancing memory of List 2. The relation of this concept to the findings in the current study will be discussed below, and its relation to the production effect will be discussed in Chapter 4.

In support of this contextual view, McElree et al. (1999) presented related evidence showing that familiarity-based processes occur prior to recollection-based processes in recognition memory. Participants studied two sets of words, of which one was read and one was heard; each set contained words that were either presented once or multiple times. On the recognition test containing words from both sets, participants were to respond only to the heard words. When relatively quick responses were required, false alarm rates were higher for the repeated read words than for the once-presented read words; however, when allowing for deliberation, false alarms for the repeated words were reduced—suggesting that participants could now remember and use the source memory of these words. This pattern of results indicates that recollection processes may not yet be available when relatively quick responses are made. Malmberg and Xu (2007) showed a similar pattern of data when participants were tested on their recognition memory of word pairs that were also studied either once or multiple times (with rearranged pairs serving as foils on the test), confirming that recollection processes require deliberation to occur and are thus unlikely to be available for participants prone to making quick responses (see also Xu & Malmberg, 2007, for formulation of a related dual-process model describing familiarity-based and recollection-based memory processes)

Sahakyan and Kelley (2002) investigated their context hypothesis using a deliberate reinstatement procedure: Prior to recalling List 1, participants who were assigned to the context reinstatement manipulation were asked to try to remember specific thoughts, feelings, or emotions that they had at the beginning of the study session and through the study phase of List 1. These could include, for example, any specific background cues that they might remember from the start of the session (e.g., the *Star Wars* theme was deliberately played), as well as any strategies that they might have used while studying List 1. There were two context change

groups—the Forget group and the Remember + Context Change group, where context change was induced between the lists by imagining walking through their parents’ house. Half of the participants in these two groups were prompted to try to remember these initial contextual cues at retrieval; half were not so prompted. Two differences—both as predicted—were observed. First, in both context change groups, memory for List 1 was improved for participants told to recollect the initial context relative to participants for whom contextual cues from List 1 were not reinstated. Second, memory for List 2 in the two context reinstatement groups was worse than for those who did not have the contextual cues from List 1 reinstated, becoming similar to that of the control Remember groups (see also Smith, 1979, 1984, for examinations of context reinstatement effects).

Rationale for the Experiment

Based on this logic, it is hypothesized that if production, as an additional dimension of encoding, serves as a contextual framework that becomes associated to produced information, prompting the use of production as a contextual cue should then facilitate memory of this produced information if it had been intentionally forgotten. In this formulation, production perhaps acts like a simple category cue which has been shown to eliminate directed forgetting effects when studying categorically related items—but only when participants are explicitly reminded of the category cue at retrieval (Lehman & Malmberg, 2011). The category cue essentially acts as an additional, specific feature/anchor which binds to the items within that category—not unlike the deliberate cues used in Sahakyan and Kelley (2002)—beyond any other background temporal-contextual cues that are not as specific (see also Lehman & Malmberg, 2009, for related modelling of such contextual-based memory effects within directed forgetting). This notion has precedent in the work of Light and Carter-Sobell (1970) who showed that

semantic adjectives can act as contextual cues to disrupt recognition memory of nouns to which these adjectives are attached.

To test this production-as-context hypothesis, a production effect manipulation was inserted into a list-method directed forgetting procedure. Participants were given two lists of words to study, both containing a mixture of items to be studied aloud and items to be studied silently. One group was instructed after studying List 1 that it would not be tested and thus did not need to be remembered (the Forget group); the other group was simply instructed to go on to study List 2 (the Remember group). If participants in the Forget group are then explicitly told to try to remember whether each item was studied aloud, this should prompt them to use the relevant contextual framework that is hypothesized to serve as a memory aid in this situation (i.e., they would be prompted to use the distinctiveness heuristic). Given this context reinstatement, memory for the aloud items in this Forget group (studied in both List 1 and List 2) should resemble that of the performance seen in the Remember group—without the typical effects of directed forgetting described above.

Recall versus Recognition

There was one additional key issue that needed to be resolved in designing the experimental procedure. In the current production effect literature, the distinctiveness account serves as an explanation for the production effect specifically when recognition memory tests are used as the method of retrieval: As discussed in Chapter 1, this specific distinctiveness account has not yet been extended into recall tests (e.g., Jonker et al., 2014, Forrin & MacLeod, 2016). However, in list-method directed forgetting, when recognition tests have been used, a directed forgetting effect often has not been observed or was only partially observed (e.g., Basden et al., 1993, Benjamin, 2006; Loft et al., 2008; Sahakyan & Delaney, 2005; Sahakyan et al., 2009).

This difference between the list-method and the item-method directed forgetting procedures has been put forth as critical evidence that these two methods for inducing intentional forgetting are caused by different cognitive mechanisms (see MacLeod, 1998, for a summary). In their original proposal, Sahakyan and Kelley (2002) presented the contextual change hypothesis as an account of directed forgetting specific to when the test is recall. In so doing, they were taking into account the apparent similarity that recognition tests often fail to find any effect not only in directed forgetting but also in (environmental) context-dependent procedures (but see Smith & Vela, 2001, for a review of when recognition tests do reveal context effects; for further modelling work examining the influence of context on recognition memory, see Dennis & Humphreys, 2001; Lehman & Malmberg, 2009; Murnane et al., 1999).

In subsequent work, Lehman and Malmberg (2009) offered a more nuanced explanation for why recognition tests generally did not show a directed forgetting effect under the list method. They argued that, according to the contextual change hypothesis, to enable recognition tests to reveal the effects of contextual change on memory in list-method directed forgetting, the recognition test procedure must be modified. At issue was whether the recognition test was one of “inclusion” versus “exclusion” (Jacoby, 1991; see also Loft et al., 2008). On an inclusion recognition test—the typically used procedure—participants only had to differentiate studied items from unstudied items (e.g., by providing a “Yes” or “No” response for each item presented at test); there was no need to differentiate whether the studied items were from List 1 or List 2. On an exclusion recognition test, however, participants would be required to remember the source list of each studied item, responding only to the items from one specified list. To do so, participants would have to differentiate the contextual cues associated with each list. Using a modified list-method procedure paired with an exclusion recognition test, Lehman and

Malmberg (2009) showed the same pattern of results as was usually seen in recall—forgetting of List 1 and facilitation of List 2 due to the mental context shift caused by the Forget instruction. In so doing, they confirmed that the exclusion format for testing recognition memory is appropriate for accurately revealing contextual differentiation effects within list-method directed forgetting.

The use of exclusion recognition tests was incorporated into the design of the present study by assigning participants to two different procedures at the time of the recognition test. Within both of the Forget and Remember groups, one subgroup was instructed to indicate for each item that they believed was studied whether that item had been studied aloud or silently (the Aloud versus Silent Differentiation condition). The other subgroup was instructed to indicate for each item that they believed was studied whether that item had been studied in List 1 or List 2 (the List 1 versus List 2 Differentiation condition). By prompting participants in the Aloud versus Silent Differentiation condition to use the distinctiveness heuristic at retrieval, the prediction was that those in this subgroup who had been instructed to forget List 1 would show a reduced effect of directed forgetting specifically for the items that had been studied aloud because they would use the relevant contextual cue that facilitates memory of the produced items. In contrast, participants in the List 1 versus List 2 Differentiation condition were expected to show the typical effects of directed forgetting for the aloud items: These participants would not have the relevant contextual cues at retrieval that would attenuate the effects of directed forgetting for these aloud items. These manipulations were not predicted to differentially affect memory for the silent items because the use of the distinctiveness heuristic represents a cue for contextual reinstatement specifically associated with the produced information (cf. Hourihan & MacLeod, 2008, who showed that in the item-method procedure of directed forgetting, a

difference also occurred whereby only the silent items—but not the produced items—were subject to directed forgetting).

In summary, the goal of the present study was to further investigate—using a converging method—whether distinctiveness within the production effect can be characterized as a context-based memory effect. A production effect manipulation was inserted within a list-method directed forgetting procedure, allowing us to investigate whether production serves as a global contextual framework that becomes associated with produced information during study. The central prediction was that despite the occurrence of a mental context change between studying two sets of information—caused by an instruction to forget—the use of the distinctiveness heuristic at retrieval would reinstate a key contextual cue that should specifically facilitate memory of produced information that was intentionally forgotten.

Method

Participants. Participants were 212 students (51 men, 153 women; age range: 17-46 years, mean age: 20.1 years, $SD = 3.7$ years; 8 participants declined to provide demographic information) from the University of Waterloo, recruited via the Department of Psychology's research participation system. Ethics approval was obtained from the University of Waterloo Research Ethics Board, and informed consent was obtained from all participants. Participants received course credit or were paid \$10 in exchange for their participation. Twenty participants were tested but excluded from analyses due to one of several factors: clear lack of effort based on evidently random responding (8 participants), difficulty in understanding instructions (5 participants), having a false alarm rate 2.5 standard deviations greater than the mean (2 participants), or technical issues (5 participants).

Of the remaining 192 participants, 95 had been assigned to the Forget group and 97 had been assigned to the Remember group. In the Forget group, 47 participants were assigned to the Aloud versus Silent Differentiation condition and 48 to the List 1 versus List 2 Differentiation condition. In the Remember group, 49 participants were assigned to the Aloud versus Silent Differentiation condition and 48 to the List 1 versus List 2 Differentiation condition.¹³

Apparatus. The experiment was initially conducted in the laboratory and was subsequently moved online due to the COVID-19 pandemic (51 participants completed the study in the laboratory, 161 participants completed the study online). In the laboratory, the study was conducted on a PC-compatible computer running a program written in E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA). Study and test trials were presented on an LCD monitor, with responses collected via a standard QWERTY keyboard. The online version of the study was conducted at Pavlovia.org running a program built in PsychoPy Version 2020.1.3 (Open Science Tools, Nottingham, England). Participants completed the experiment on their own computers. An experimenter monitored participants for the duration of the study via online video using Cisco Webex (Cisco Systems, San Jose, CA) to ensure understanding and compliance with all of the instructions.

¹³ A post-hoc power analysis was performed using G*Power Version 3.1.9.4 (Faul et al., 2007) to confirm that for the aloud items studied in the Aloud versus Silent Differentiation condition, if retrieval using production as a contextual cue did not sufficiently offset the typically expected effects from the directed forgetting manipulation (which are expected in the List 1 versus List 2 Differentiation condition), that a sufficient number of participants had been included to detect these effects. The main hypothesis of this study was, of course, that reinstating the use of production as a contextual cue at retrieval would offset the typically expected effects from receiving an instruction to forget. The effect size used in this calculation was based on that from the 2 (List 1/List 2; within) \times 2 (Forget/Remember; between) mixed ANOVA for the aloud items in the List 1 versus List 2 Differentiation condition ($\eta_p^2 = 0.120$). The reasoning was that if the main hypothesis was incorrect, a similar difference between the Forget and Remember groups should be seen between the two experimental conditions. Based on $\alpha = .05$ and power = 0.95, the appropriate sample size needed to satisfy these parameters was estimated to be $N = 34$ for the Forget and Remember groups combined. Therefore, the sample size of $N = 96$ is more than adequate for investigating this main hypothesis in the current experiment.

Materials. The study materials consisted of the set of 240 words listed in Appendix B. For the laboratory group, the words were presented in lowercase in the Consolas font, size 36; for the online group, the words were presented with a size of 0.05 height units (i.e., a height of 5% of the screen height). In both cases, the words were presented against a black background. Words assigned to each condition (i.e., aloud, silent, or new) and to each list were selected randomly for each participant. Participants studied 40 words aloud and 40 words silently in each of their two lists, with the sequence of words and instructional conditions randomized. The remaining 80 words were used as distractors on the recognition test where the resulting 240 test words were randomized anew.

Procedure. The procedure is illustrated in Figure 2. For the study phase, participants were instructed that they would be studying two lists of words. For both lists of words, they were to read the words presented in blue aloud and the words presented in white silently; silent reading was to be done without moving their lips. Each blue or white study word was presented individually for 3 s at the centre of the screen, and a blank period of 500 ms intervened between successive stimuli. After studying List 1, participants in the Forget group were told that only List 2 would be tested, and they should proceed to study List 2 in the same manner as List 1 (i.e., reading words aloud or silently based on colour). Participants in the Remember group were simply instructed to continue on to study List 2 after completing List 1.

A recognition test immediately followed the study phase, with all participants tested on both lists of studied words. The 160 studied words, intermingled with the 80 distractor words, were presented one at a time in a random order in yellow, so that colour would not serve as a retrieval cue. In the Aloud versus Silent Differentiation condition, participants used a key press to indicate whether each word was studied aloud, was studied silently, or was new (the “A” key

for aloud; the “L” key for silent, and the “N” key for new). In the List 1 versus List 2 Differentiation condition, participants used a key press to indicate whether each word was studied in List 1, was studied in List 2, or was new (the “1” key for List 1; the “2” key for List 2, and the “N” key for new). Participants could take as long as they needed for each response. Upon response, the word disappeared and, after a 500-ms blank screen, the next word appeared. The entire procedure took under 30 minutes.

Figure 2

The study procedure. All participants studied two lists of words, half of each list aloud and half silently. After studying List 1, and prior to studying List 2, the Forget group was instructed that List 1 would not be tested so it did not need to be remembered; the Remember group was simply told to continue. After studying List 2, all participants were tested on both lists, either by indicating whether each word was studied in List 1 or in List 2, or by indicating whether each word was studied aloud or silently.



The procedure for the online version was the same as for the laboratory version, except that participants began by meeting an experimenter on the Cisco Webex online video platform, and the experimenter then provided them with the study's web link. Due to the extra time required for setup, the online version of the study typically took slightly longer to complete than the time required in the laboratory but not more than 60 minutes in total.

Results

The analyses were conducted in JASP Version 0.14 (JASP Team, 2020). The recognition test data are shown in Tables 2 and 3 and corresponding Figures 3 and 4. Analyses first examined overall performance (Table 2 and Figure 3). False alarm rates were low and did not differ across the four groups, $F(3, 188) = 2.14$, $MSE = 0.030$, $p = .097$, $\eta_p^2 = 0.033$. That false alarm rates were lower than hit rates is evidence of relatively good memory for the studied words.

A series of ANOVAs was then performed to examine memory for the studied items as a function of differentiation condition: (1) Aloud versus Silent or (2) List 1 versus List 2.

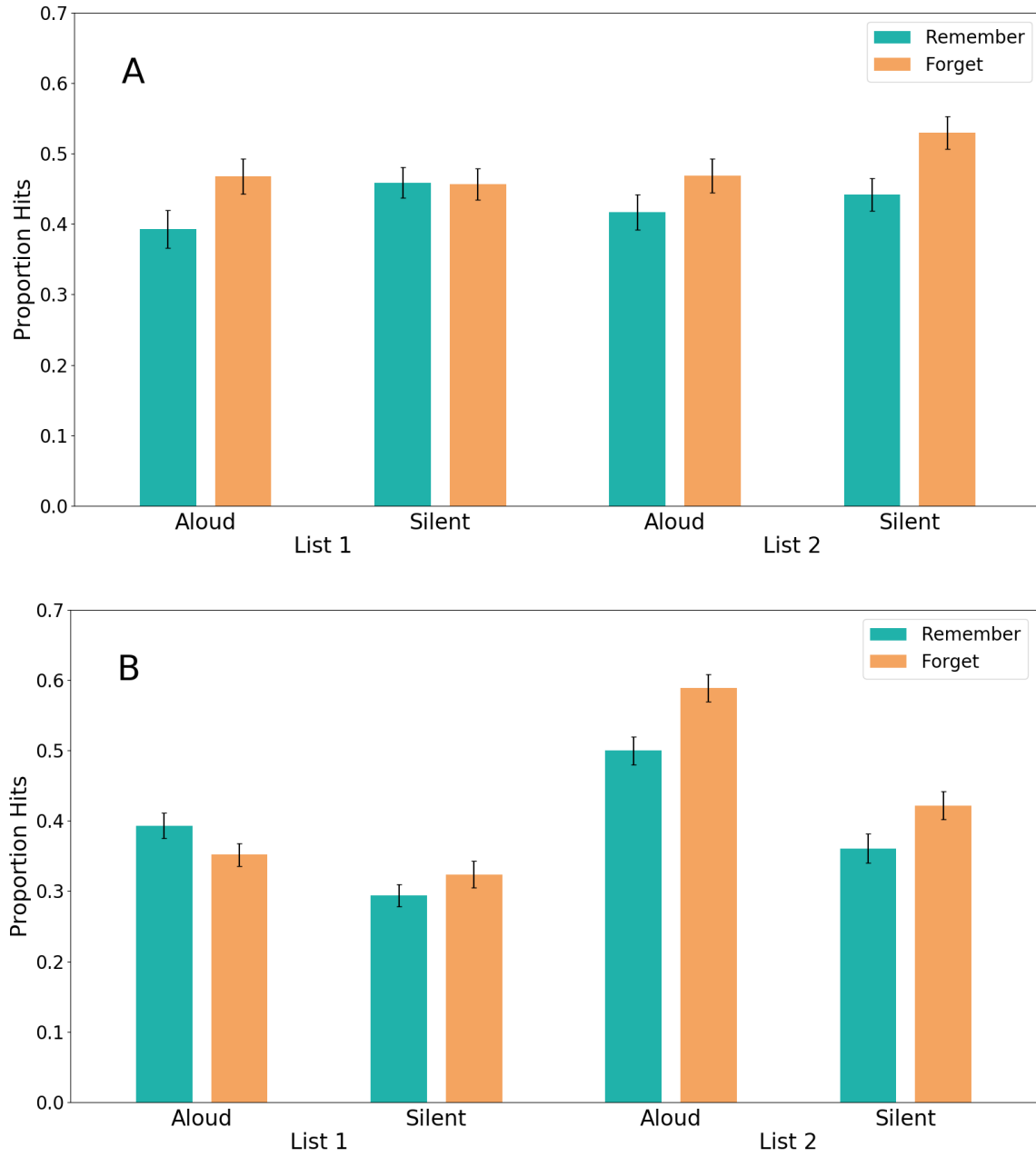
Table 2*Proportions of hits for studied targets and false alarms for unstudied distractors.*

Study condition	List 1		List 2		Distractor
	Aloud	Silent	Aloud	Silent	
<i>Aloud vs. silent Differentiation</i>					
Remember	.393 (.027)	.459 (.022)	.417 (.025)	.442 (.023)	.276 (.020)
Forget	.468 (.025)	.457 (.022)	.469 (.024)	.530 (.023)	.338 (.026)
<i>List 1 vs. List 2 Differentiation</i>					
Remember	.393 (.018)	.294 (.016)	.500 (.020)	.361 (.021)	.357 (.028)
Forget	.352 (.016)	.324 (.019)	.589 (.019)	.422 (.020)	.301 (.025)

Note. Standard errors shown in parentheses below each mean.

Figure 3

Proportions of hits for studied targets.



Note. **A:** Aloud versus Silent Differentiation condition; **B:** List 1 versus List 2 Differentiation condition. Error bars represent standard error.

Aloud versus Silent Differentiation. A 2 (List 1/List 2; within) \times 2 (Aloud/Silent; within) \times 2 (Forget/Remember; between) mixed ANOVA was performed first to test the prediction that the influence of the directed forgetting manipulation would differ depending on whether the aloud or the silent items were examined. The Forget group showed better memory overall than the Remember group, $F(1, 94) = 5.03$, $MSE = 0.054$, $p = .027$, $\eta_p^2 = 0.051$. Memory did not differ between lists, $F(1, 94) = 3.91$, $MSE = 0.010$, $p = .051$, $\eta_p^2 = 0.040$, nor between aloud versus silent study, $F(1, 94) = 3.09$, $MSE = 0.039$, $p = .082$, $\eta_p^2 = 0.032$. None of the three 2-way interactions was significant: 2 (List 1/List 2; within) \times 2 (Aloud/Silent; within): $F(1, 94) = 0.86$, $MSE = 0.007$, $p = .355$, $\eta_p^2 = 0.009$; 2 (List 1/List 2; within) \times 2 (Forget/Remember; between): $F(1, 94) = 2.73$, $MSE = 0.010$, $p = .102$, $\eta_p^2 = 0.028$; 2 (Aloud/Silent; within) \times 2 (Forget/Remember; between): $F(1, 94) = 0.25$, $MSE = 0.039$, $p = .620$, $\eta_p^2 = 0.003$.

The critical 3-way interaction was significant, $F(1, 94) = 11.22$, $MSE = 0.007$, $p = .001$, $\eta_p^2 = 0.107$, confirming the prediction that aloud versus silent study is a contributing factor toward differences in recognition memory resulting from the context reinstatement manipulation. A pair of 2 (List 1/List 2; within) \times 2 (Forget/Remember; between) mixed ANOVAs were thus performed to examine the hit rates as a function of list and study instruction group, separately for the aloud and silent items. Consider first the aloud items. Memory did not differ between lists, $F(1, 94) = 0.84$, $MSE = 0.009$, $p = .361$, $\eta_p^2 = 0.009$. Although participants in the Forget group showed numerically better overall recognition than those in the Remember group, this instructional difference was only marginal statistically, $F(1, 94) = 3.64$, $MSE = 0.053$, $p = .060$, $\eta_p^2 = 0.037$. The interaction was not significant, $F(1, 94) = 0.64$, $MSE = 0.009$, $p = .425$, $\eta_p^2 =$

0.007.¹⁴ Next consider the silent items. A corresponding ANOVA showed that recognition was overall better for List 2 than for List 1, $F(1, 94) = 4.82$, $MSE = 0.008$, $p = .031$, $\eta_p^2 = 0.049$, but that memory did not differ as a function of instruction group, $F(1, 94) = 2.20$, $MSE = 0.040$, $p = .141$, $\eta_p^2 = 0.023$. Because the interaction was significant, $F(1, 94) = 12.35$, $MSE = 0.008$, $p < .001$, $\eta_p^2 = 0.116$, the relevant simple effects were probed. Paired-samples t -tests showed that there was no difference in memory between the Forget and Remember groups for List 1, $t(94) = 0.06$, $p = .956$, $d = 0.011$, but that there was significantly better memory in the Forget group than in the Remember group for List 2, $t(94) = 2.73$, $p = .008$, $d = 0.557$. This was entirely due to better memory in List 2 relative to List 1 in the Forget group, $t(46) = 3.48$, $p = .001$, $d = 0.507$; there was no significant difference in memory between Lists 1 and 2 in the Remember group, $t(48) = 1.13$, $p = .263$, $d = 0.162$.

Thus, for the aloud items, no effect of directed forgetting was observed on recognition memory in the Aloud versus Silent Differentiation condition. This is consistent with the prediction that production serves as a contextual framework which can later facilitate memory of produced information when this information had been intentionally forgotten—but only if people use this contextual cue at retrieval. For the silent items, memory for the List 2 items was better in the Forget groups than in the Remember groups, consistent with the expectation. However, memory for the List 1 items was similar between the Forget and the Remember groups.

List 1 versus List 2 Differentiation. The same 3-way mixed ANOVA was performed first to test the prediction that the influence of the directed forgetting manipulations would differ for

¹⁴ A supporting Bayesian repeated measures ANOVA, conducted using uniform priors, revealed a Bayes factor indicating moderate evidence for the null hypothesis (Lee & Wagenmakers, 2013) for the interaction, $BF_{01}=3.704$. (Note that the error percentage from this analysis was 4.716%, indicating that the Bayes factor will change slightly when the analysis is repeated; however, this is within the acceptable error percentage limit: See van Doorn et al., 2020.)

the aloud items versus the silent items. List 2 was overall better remembered than List 1, $F(1, 94) = 75.11$, $MSE = 0.021$, $p < .001$, $\eta_p^2 = 0.444$, and aloud items showed better memory overall than silent items, $F(1, 94) = 154.16$, $MSE = 0.007$, $p < .001$, $\eta_p^2 = 0.621$. The Forget group also showed a small memory advantage overall than the Remember group, $F(1, 94) = 4.00$, $MSE = 0.029$, $p = .049$, $\eta_p^2 = 0.041$. A significant 2 (List 1/List 2; within) \times 2 (Forget/Remember; between) interaction showed that this memory advantage was primarily due to better memory in the Forget group for List 2, $F(1, 94) = 7.61$, $MSE = 0.021$, $p = .007$, $\eta_p^2 = 0.075$. In addition, a significant 2 (List 1/List 2; within) \times 2 (Aloud/Silent; within) interaction revealed that the better memory for aloud over silent items was greater in List 2 than in List 1, $F(1, 94) = 19.85$, $MSE = 0.010$, $p < .001$, $\eta_p^2 = 0.174$. The 2 (Aloud/Silent; within) \times 2 (Forget/Remember; between) interaction was not significant, $F(1, 94) = 1.62$, $MSE = 0.007$, $p = .206$, $\eta_p^2 = 0.017$.

The critical 3-way interaction was significant, $F(1, 94) = 6.08$, $MSE = 0.010$, $p = .015$, $\eta_p^2 = 0.061$, confirming the prediction that aloud versus silent study is a contributing factor toward differences in recognition memory resulting from the context reinstatement manipulation. As was done for the Aloud versus Silent differentiation, a pair of 2 (List 1/List 2; within) \times 2 (Forget/Remember; between) mixed ANOVAs was then performed to examine the hit rates as a function of list and study instruction group, separately for the aloud and silent items. Consider first the aloud items. Recognition was significantly better for List 2 items than for List 1 items, $F(1, 94) = 88.51$, $MSE = 0.016$, $p < .001$, $\eta_p^2 = 0.485$, but recognition did not differ between the two instructional groups, $F(1, 94) = 1.73$, $MSE = 0.016$, $p = .191$, $\eta_p^2 = 0.018$. Because the interaction was significant, $F(1, 94) = 12.76$, $MSE = 0.016$, $p < .001$, $\eta_p^2 = 0.120$, the relevant simple effects were probed. Paired-samples t -tests showed significantly worse memory in List 1 for the Forget group than for the Remember group, $t(94) = 1.74$, $p = .043$, $d = 0.354$; this pattern

reversed in List 2 where memory was significantly better in the Forget group than in the Remember group, $t(94) = 3.27, p < .001, d = 0.668$. Now consider the silent items. A corresponding ANOVA showed that recognition was again significantly better for List 2 than for List 1, $F(1, 94) = 22.61, MSE = 0.014, p < .001, \eta_p^2 = 0.194$, and was also significantly better in the Forget group than in the Remember group, $F(1, 94) = 4.86, MSE = 0.021, p = .030, \eta_p^2 = 0.049$.¹⁵ The interaction was not significant, $F(1, 94) = 0.82, MSE = 0.014, p = .369, \eta_p^2 = 0.009$.

In the List 1 versus List 2 Differentiation condition, the expected pattern of List 1 forgetting and List 2 facilitation was thus observed, whereas this pattern was not observed in the Aloud versus Silent Differentiation condition. Again, these results agreed with the prediction that production is a contextual framework that can facilitate memory of to-be-forgotten information at retrieval when that information had been produced during study. Attempting to vaguely differentiate the studied information by list (i.e., without specific contextual anchors) does not reinstate the critical contextual cues that would enhance memory of information intentionally forgotten, instead leading to the typical effects from directed forgetting.

For the silent items in both experimental conditions, memory for the List 2 items was better in the Forget groups than in the Remember groups, consistent with expectation.

Unexpectedly, though, memory for the List 1 items was similar between the Forget and the Remember groups. These results are further elaborated in the Discussion.

Next, corresponding analyses were performed for the target items that were correctly remembered but that participants attributed to the wrong source (Table 3 and Figure 4). The goal here was to show converging evidence that if List 2 was indeed differentiated from List 1 due to

¹⁵ A post-hoc analysis showed, however, that although memory for the List 1 silent items was numerically better in the Forget group than in the Remember group for this condition, there was no difference in memory statistically, $t(94) = 1.22, p = .225, d = 0.249$.

a mental context change, fewer errors should be made when remembering List 2 than List 1 because the context associated with List 2 will be more similar to that at retrieval than will the context associated with List 1. This is precisely the result expected for the aloud items studied in the List 1 versus List 2 Differentiation condition because this manipulation does not reinstate the critical contextual framework of production at retrieval for List 1. In contrast, no difference is expected in the Aloud versus Silent Differentiation condition because reinstatement of context is expected to overcome the contextual differentiation between List 1 and List 2 caused by the Forget instruction.

Table 3

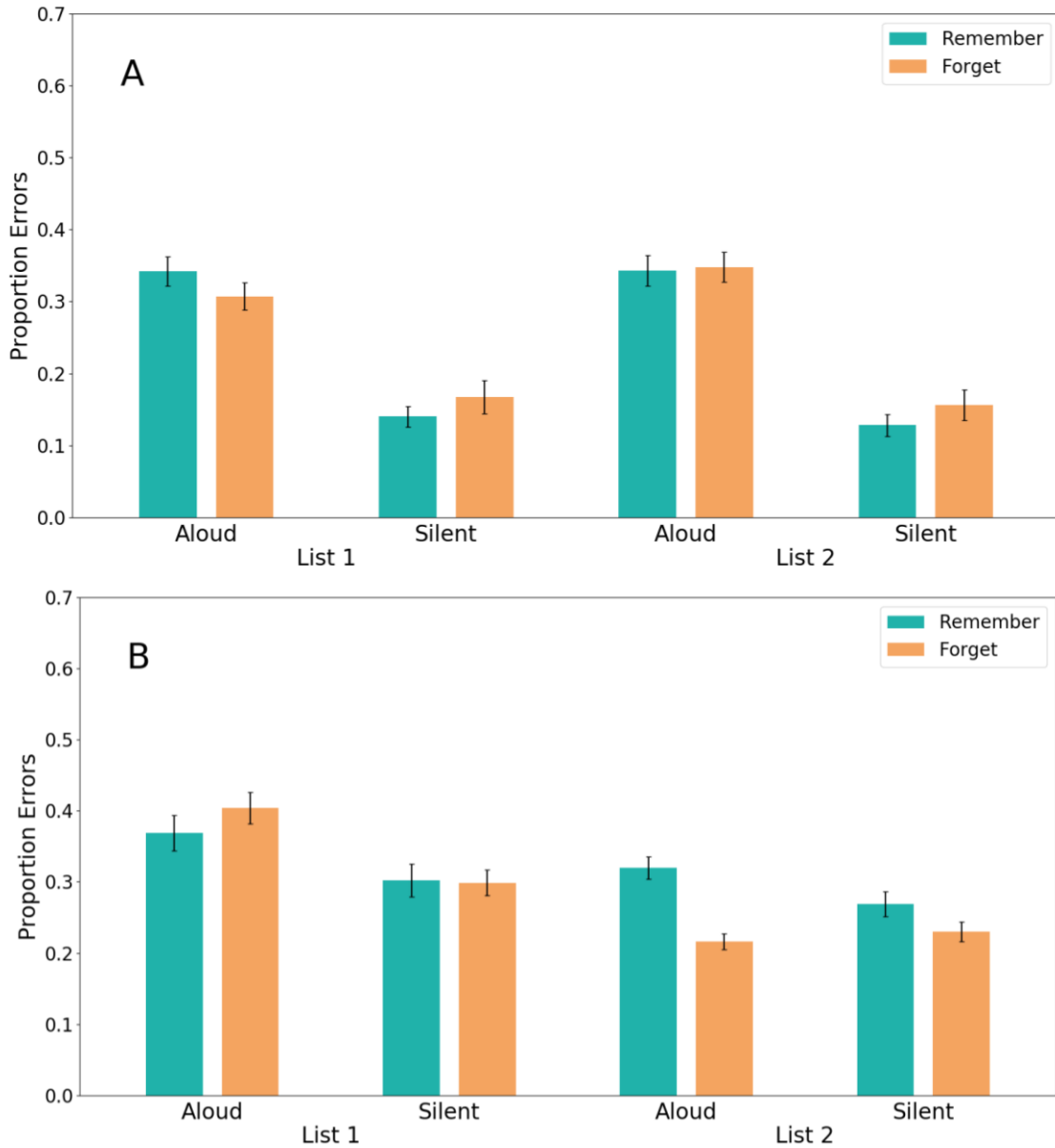
Proportions of studied targets that were correctly recognized but that were associated with the incorrect source (e.g., indicating an item studied aloud as an item studied silently; indicating an item studied in List 1 as an item studied in List 2).

Study condition	List 1		List 2	
	Aloud	Silent	Aloud	Silent
<i>Aloud vs. silent Differentiation</i>				
Remember	.342 (.020)	.140 (.014)	.343 (.021)	.128 (.015)
Forget	.307 (.019)	.167 (.023)	.348 (.021)	.156 (.021)
<i>List 1 vs. List 2 Differentiation</i>				
Remember	.369 (.025)	.302 (.023)	.320 (.016)	.269 (.018)
Forget	.404 (.022)	.299 (.018)	.216 (.011)	.230 (.014)

Note. Standard errors shown in parentheses below each mean.

Figure 4

Proportions of studied targets that were correctly recognized but that were associated with the incorrect source.



Note. **A:** Aloud versus Silent Differentiation condition (indicating an item studied aloud as an item studied silently, or vice versa); **B:** List 1 versus List 2 Differentiation condition (indicating an item studied in List 1 as an item studied in List 2, or vice versa). Error bars represent standard error.

Aloud versus Silent Differentiation. Source errors sometimes occurred such that items that were studied aloud (in either list) were indicated by participants to have been studied silently, or vice versa. Two 2 (List 1/List 2; within) \times 2 (Forget/Remember; between) mixed ANOVAs were performed again, separately for aloud and silent items, to examine the rates of source judgment errors as a function of list and instructional group. There were no significant main effects nor interactions—for Aloud items: no main effect of list, $F(1, 94) = 3.00$, $MSE = 0.007$, $p = .087$, $\eta_p^2 = 0.031$, no main effect of group, $F(1, 94) = 0.33$, $MSE = 0.033$, $p = .568$, $\eta_p^2 = 0.003$, and no interaction, $F(1, 94) = 2.71$, $MSE = 0.007$, $p = .103$, $\eta_p^2 = 0.028$; for Silent items: no main effect of list, $F(1, 94) = 2.03$, $MSE = 0.003$, $p = .157$, $\eta_p^2 = 0.021$, no main effect of group, $F(1, 94) = 1.22$, $MSE = 0.029$, $p = .273$, $\eta_p^2 = 0.013$, and no interaction, $F(1, 94) < 0.01$, $MSE = 0.003$, $p = .948$, $\eta_p^2 < 0.001$. The results confirm the above prediction for the aloud items in this condition—that there would be no difference in the number of errors between the Forget and Remember groups because production, as the critical contextual framework during encoding, is highlighted again at retrieval.

List 1 versus List 2 Differentiation. Source errors sometimes also occurred such that items that were studied in List 1 (either aloud or silently) were indicated by participants to have been studied in List 2, or vice versa. Two 2 (List 1/List 2; within) \times 2 (Forget/Remember; between) mixed ANOVAs were performed, separately for aloud and silent items, to examine the rates of source judgment errors as a function of list and instructional group. For items studied aloud, participants made significantly more source judgment errors for items studied in List 1 than in List 2, $F(1, 94) = 36.39$, $MSE = 0.018$, $p < .001$, $\eta_p^2 = 0.279$, but these source errors did not differ between groups, $F(1, 94) = 3.52$, $MSE = 0.017$, $p = .064$, $\eta_p^2 = 0.036$. The interaction was significant, $F(1, 94) = 12.68$, $MSE = 0.018$, $p < .001$, $\eta_p^2 = 0.119$, so the relevant simple effects

were probed. Paired-samples t -tests showed that, for aloud items studied in List 1, errors did not differ between the Forget and the Remember groups, $t(94) = 1.06$, $p = .291$, $d = 0.217$ (though numerically, participants in the Forget group made more errors than the Remember group); in contrast, for aloud items studied in List 2, errors were significantly higher in the Remember group than in the Forget group, $t(94) = 5.38$, $p < .001$, $d = 1.098$. Apparently, it was more difficult for participants to distinguish the source for the list 2 aloud items when they had been instructed to remember List 1 relative to when they had been instructed to forget List 1, suggesting that participants in the Forget group experienced a reset of contextual sampling that enabled greater distinction between Lists 1 and 2 at retrieval—consistent with the earlier prediction. For items studied silently, participants also made significantly more source judgment errors for items studied in List 1 than for items studied in List 2, $F(1, 94) = 9.69$, $MSE = 0.013$, $p = .002$, $\eta_p^2 = 0.093$, possibly due to believing that remembered items were more likely to be more recently learned. Errors did not differ between the instructional groups, $F(1, 94) = 1.08$, $MSE = 0.019$, $p = .301$, $\eta_p^2 = 0.011$, and the interaction was not significant, $F(1, 94) = 1.21$, $MSE = 0.013$, $p = .274$, $\eta_p^2 = 0.013$.

Discussion

The goal of this study was to further examine whether the production effect observed in recognition memory could be characterized as a context-based memory effect. To accomplish this, a production manipulation was inserted within a list-method directed forgetting procedure. If production can act as a contextual framework to which produced information becomes associated then, at retrieval, remembering whether studied items were produced (i.e., using the distinctiveness heuristic) should improve memory of the produced information that had been intentionally forgotten. According to the contextual change hypothesis of list-method directed

forgetting (Sahakyan & Kelley, 2002), forgetting occurs because people experience a mental context change after being instructed that they do not need to remember information studied earlier. Later reinstating the relevant contextual framework with which this information is associated will facilitate memory of the forgotten information.

Two experimental conditions were thus created, both of which tested participants on their recognition memory for the studied items but each of which required a different differentiation. In the Aloud versus Silent Differentiation condition, participants responded according to whether the studied items were aloud or silent: This is the condition in which the relevant contextual cue—production—should be reinstated and should thus attenuate the expected effects of directed forgetting on the items that were studied aloud in List 1. In the List 1 versus List 2 Differentiation condition, participants responded according to whether the studied items were from List 1 or List 2: As this procedure does not prompt the use of production as the relevant contextual cue from List 1, the typical pattern of directed forgetting effects was expected in this condition.

The results supported the predictions for the aloud items. In the Aloud versus Silent Differentiation condition, there was no difference in memory between the Forget and the Remember groups. Thus, the manipulation of explicitly prompting participants to use the distinctiveness heuristic at retrieval allowed participants to overcome the otherwise expected forgetting of the List 1 items. With respect to the contextual change hypothesis, this result suggests that production indeed represents a contextual framework with which produced items become associated. This framework then facilitates memory of the intentionally forgotten produced items when participants are prompted to use it at retrieval. These List 1 produced items

would otherwise suffer from forgetting because of a mental contextual change from being instructed to forget List 1 prior to studying List 2.

The production effect observed in recognition memory can thus be conceptualized as a context-based memory effect and the distinctiveness heuristic as a technique that uses an associative contextual cue to facilitate memory of produced information. In further support of this conclusion, there were no differences in the source judgment errors between the Forget and Remember groups in the Aloud versus Silent Differentiation condition (i.e., incorrectly indicating that an aloud item was a silent item, or vice versa). This suggests that the use of the distinctiveness heuristic allowed participants in the Forget group to essentially rejoin the two studied lists into one list to be remembered. Otherwise, when retrieving List 2, the Forget group would be expected to make fewer errors than the Remember group because context would remain the same with only that of List 2 at retrieval—participants in the Forget group would be better able to focus on List 2 because they were instructed to forget List 1.

In contrast, for the List 1 versus List 2 Differentiation condition, the expected pattern of directed forgetting effects was found for the aloud items: In the Forget group, participants showed worse memory for List 1 and better memory for List 2 compared to the Remember group. With respect to the contextual change hypothesis, this is the pattern of results that would be predicted to occur when the relevant contextual cues are not reinstated at retrieval for the Forget group. Whereas invoking the distinctiveness heuristic in the Aloud versus Silent Differentiation condition served as an associative contextual cue that facilitated memory of the intentionally forgotten List 1 aloud items, in the List 1 versus List 2 Differentiation condition, context remained similar to that of List 2 at retrieval because list was not a distinct contextual cue, resulting in forgetting of the List 1 items but a memory enhancement for the List 2 items.

Moreover, compared to the Remember group, participants in the Forget group were less likely to make source judgment errors for aloud items studied in List 2 (i.e., incorrectly indicating that a List 2 item was studied in List 1): By inducing a context change with a Forget instruction and not reinstating the relevant contextual cues at retrieval, participants effectively had a smaller pool of information to attend to and were hence less likely to make errors for the more recently studied List 2 items.

A different pattern emerged for the silent items. In both the Aloud versus Silent Differentiation condition and the List 1 versus List 2 Differentiation conditions, the silent items from List 2 were better remembered by the Forget group than by the Remember group, whereas memory for the List 1 silent items did not differ between the Forget and the Remember groups. With respect to the contextual change hypothesis, better memory for the List 2 items in the Forget groups compared to the Remember groups is expected, given that a change in mental context would have occurred between Lists 1 and 2 in the Forget groups and the experimental manipulations did not explicitly reinstate any relevant contextual cues from List 1 for the silent items at retrieval, resulting in a context mismatch between List 1 and retrieval. However, why was memory for the List 1 silent items similar across the Forget and Remember groups within both retrieval instruction conditions?

Pastotter et al. (2012; see also Pastotter & Bauml, 2010, Pastotter et al., 2017) proposed a reset-of-encoding hypothesis as an alternative explanation for the memory enhancement of List 2 information in list-method directed forgetting. Essentially, providing a Forget instruction after List 1 “resets” working memory load and thereby increases the efficacy of the encoding of (at least the initial items of) List 2 compared to the Remember condition. Using serial position analyses, initial studies (Pastotter & Bauml, 2010; Pastotter et al., 2012) confirmed this

prediction of the memory enhancement for List 2 when participants were tested via recall. In addition, Pastotter et al. (2016) showed a similar List 2 memory enhancement effect as well as a lower false alarm rate for List 2 in the Forget group (by testing Lists 1 and 2 separately) when examining participants' recognition memory. Notably, the expected effect of List 1 forgetting was absent, as has been the case in some previous directed forgetting studies that had examined recognition memory (e.g., Benjamin, 2006; Sahakyan & Delaney, 2005). The reset-of-encoding hypothesis proposes a dual-mechanism view of list-method directed forgetting (see also Sahakyan & Delaney, 2005) where the reset of working memory following the Forget instruction provides only an effect of improved encoding efficacy for List 2, while a different mechanism explains forgetting effects (or lack thereof) in List 1 as well as possibly further contributing to the memory facilitation effects in List 2.

In the current study, it is possible that the silent items were perceived as “background” items relative to the distinct aloud items. Indeed, this would be consistent with the idea of distinctiveness. As discussed in Chapter 1, Hunt (2013, p. 10) defines distinctive processing as “the processing of difference in the context of similarity”. In this study, all of the target items were similar in that they were all words being presented in the same font on a computer screen, but the aloud items were made more distinct by production serving as an additional encoding dimension. It is thus possible that due to its distinctiveness, production served as the most obvious contextual framework in the present experimental procedure, which primarily affected the aloud items: Any contextual elements associated with the silent items would have been less apparent to the participants. Consequently, contextual change would have had little effect on memory for the silent items, whereas a memory facilitation effect for the List 2 silent items in the

Forget groups would still have been expected due to reset of encoding as a result of the instruction to forget List 1.

Overall, then, the results can be seen as fitting a dual-mechanism explanation of directed forgetting: Reset of encoding would have resulted in a memory enhancement in List 2 for both the aloud and the silent items whereas, specific to the experimental manipulations, List 1 forgetting and List 2 memory enhancement due to contextual effects only affected the aloud items. More generally, then, the various mechanisms that contribute to directed forgetting can differentially affect memory of the target information when a mixture of methods is used to study that information.

As an alternative to the reset-of-encoding hypothesis, Sahakyan and Delaney (2003) proposed that the use of a more effective strategy for studying List 2 as a result of receiving a Forget instruction for List 1 could account for the memory enhancement of List 2 in the Forget group (although the goal was to avoid this type of effect—see below). This would be a similar dual-mechanism view for explaining the one-sided List 2 memory enhancement effects for the silent items: These silent items benefit from a better encoding strategy compared to List 1, while the contributions of contextual effects to memory for the aloud items remain as a relevant but separate mechanism. Future studies should continue to investigate whether multiple mechanisms in combination contribute to list method directed forgetting effects as a function of how information is studied.

There are two further considerations for future work. First, Abel and Bauml (2019) examined the effect of temporal delays between study and retrieval in list-method directed forgetting. The effects of short versus long delays differed between the directed forgetting conditions and the conditions where a mental context change was induced simply via an

imagination task: Whereas directed forgetting effects (focused on List 1 recall) persisted regardless of the retrieval delay, overall memory for the studied information was reduced with the long delay but memory was maintained in the imagination conditions. Abel and Bauml (2019) suggested that their finding dissociated directed forgetting and forgetting induced by explicit context changes. However, this does not account for the possibility that directed forgetting effects are due to a combination of contextual change as well as the instruction to intentionally forget studied information. In addition, their study did not further investigate possible effects of context reinstatement and also only examined recall but not recognition memory. Future studies should continue to examine the influence of temporal delays on retrieval in directed forgetting, particularly in terms of context reinstatement effects (cf. Shin et al., 2021, showing that temporal delays between study and test did not influence context reinstatement effects on recall memory).

With regard to the second consideration, the method of study (i.e., reading words aloud or silently) was maintained between List 1 and List 2, since typical list-method directed forgetting studies do not explicitly specify changes in encoding strategies from List 1 to List 2, and because it was also necessary to avoid confounding any effects from contextual change with changes in encoding strategy (Sahakyan & Delaney, 2003). However, this decision may have affected the actual amount of contextual change experienced by the participants, as the study phase of both lists would have had production as an associative contextual cue. Future work should investigate whether changing the encoding method after a Forget instruction (i.e., for List 2) influences the benefit of the later reinstatement of List 1 contextual cues.

In summary, this study provides further evidence that the production effect observed in recognition memory can be characterized as a context-based memory effect. When participants

were prompted to use the distinctiveness heuristic at retrieval, production—as the relevant contextual framework to which the List 1 aloud items were associated—was reinstated, facilitating memory for the intentionally forgotten information. In contrast, using list as a contextual cue did not change the typically expected directed forgetting effects. With respect to the silent items, the one-sided memory enhancement of List 2 without forgetting effects in List 1 supported a dual-mechanism view of directed forgetting, with the lack of forgetting effects for the List 1 silent items attributed to reduced contextual effects due to the possibility of the silent items being situated in the background against the distinct aloud items. Taken together, the findings provide additional support for a context-based view of the production effect.

Chapter 4 – The Production Effect in Recall

This chapter will cover Study 3: The examination of how the distinctiveness of aloud items, as a contextual-based effect, may interfere with the study of other items (aloud or silent) when the test is recall. To preface, the goal of this study is to provide a more concrete conceptualization of the notion of distinctive or “unusual” processing of produced information in the situation of recall. In line with the studies in the preceding chapters of the dissertation, this will be accomplished by empirically considering the production effect in recall to be a context-based memory effect. Specifically, in recall, might the distinctive processing afforded by the produced items represent a global contextual cue that becomes associated with (only) these items during encoding, which consequently interferes with the retention of the relational information for the unproduced items due to a release from buildup of proactive interference?

Task Switching in Mixed-List Designs

To address this question, the first requirement was an experimental design that would permit removal of any possible effects of task switching during encoding from within the mixed-list production effect procedure. Such an influence within the production effect was proposed by Lambert et al. (2016), who failed to find that participants used item-order memory to support their recall (although in their study participants had to learn relatively long lists of 48 items). Lambert et al. (2016) proposed that apart from the disruption of relational memory, studying a series of items in multiple, randomized ways, as in the mixed-list production effect procedure, could induce an overall cost due to task switching. This task-switching cost would impair learning of both the aloud items and the silent items; however, the distinctive encoding afforded by the aloud items would offset the cost for those items, and thus a cost of task switching would be evident only for the silent items.

In the present study, a modified mixed-list design was used such that each participant still studied both aloud and silent items, but these items were now studied separately from each other. This paradigm required that each subject study multiple pairs of lists: The specific arrangement of “trials” is shown in Table 4. The design was essentially a $2 \times 2 \times 2$ within-subject design. List 1 could be studied either aloud or silently, List 2 could be studied either aloud or silently, and the test could be recall of either List 1 or List 2. Recall of List 1 permitted measurement of the retroactive interference due to List 2; recall of List 2 permitted measurement of the proactive interference due to List 1. Two additional single (i.e., pure) list conditions were also included as baselines: Aloud only and silent only—which were studied by the same group of participants.

Table 4

List of the ten study trials in Study 3 (presented in random order).

Trial	Studied lists		Tested list	Type of interference
	List 1	List 2		
1	Aloud	Silent	List 1	Retroactive
2	Aloud	Silent	List 2	Proactive
3	Silent	Aloud	List 1	Retroactive
4	Silent	Aloud	List 2	Proactive
5	Aloud	Aloud	List 1	Retroactive
6	Aloud	Aloud	List 2	Proactive
7	Silent	Silent	List 1	Retroactive
8	Silent	Silent	List 2	Proactive
9	Aloud	-	List 1	Aloud baseline
10	Silent	-	List 1	Silent baseline

As shown in the first four rows of Table 4, for these trials, participants studied one list of items aloud and one list of items silently, or vice versa, in succession, and were tested on their recall of either List 1 or List 2. Performance on these trials was compared with that of the two single trials (the last two rows). In line with the item-order account, a production effect was expected when the silent list was studied prior to the aloud list due to the disruption of relational processing by the aloud list, resulting in a memory cost for the silent list. In contrast, when the aloud items were studied prior to the silent items, no production effect would be expected: Relational processing of the silent list should not be disrupted and should therefore offset the benefit of production for the aloud list.

Context-Based Effects in Recall

Returning now to the main goal of characterizing the item-order account of the production effect as constituting a context-based memory effect, two ways in which context effects of production would be revealed can be foreseen in the modified mixed-list design. First, relative to the single-list trials in the current design, the aloud and the silent items in the dual-list trials should both proactively and retroactively interfere with the memory of each other (i.e., memory for both aloud and silent items is expected to be lower in the dual-list relative to the single-list trials), due to noisy temporal-contextual cues between lists affecting recall (Unsworth et al., 2012, 2013). Essentially, because the two lists are studied in close temporal proximity in the dual-list trials, participants will experience uncertainty when attempting to reconstruct the contextual cues associated with only the single target list at recall. They will, as result, incorrectly recall some items from the other list due to setting a more liberal criterion in their mental search set to try to ensure that target items are not missed.

As discussed in Chapter 1, existing theoretical models examining effects of contextual cueing at recall (e.g., Howard & Kahana, 2002; Mensink & Raaijmakers, 1988) suggest that, as a memory aid, people encode the active contextual cues at study and then attempt to use these contextual cues to remember the associated studied information at retrieval. This Noisy Context Hypothesis provides an extension by suggesting that an imperfect mechanism of reconstructing the context associated with previously studied information is used when there is uncertainty regarding which contextual cues are relevant (e.g., when different information is encoded successively), which needs to be incorporated into the extant theories.

For two sets of information both studied in the same, non-distinct way (e.g., both via silent reading), the noisy reconstruction of context would be expected to similarly affect the recall of both sets (Unsworth et al., 2013). However, when aloud items are studied after silent items, it is hypothesized that the distinctive encoding of the aloud items (i.e., by having a distinct contextual cue) would reduce the amount of interference on these items due to a release from buildup of proactive interference (see Kliegl & Bauml, 2021, for a review).

Proactive interference—the disruptive influence of previously studied information on the memory of new information—has been argued to be one of the main causes of forgetting (e.g., Anderson & Neely, 1996; Crowder, 1976; Underwood, 1957). Memory research has demonstrated, though, that a range of techniques related to contextual cueing can be used to overcome the cost due to proactive interference. For example, interpolated testing has been demonstrated to invoke mental context shifts between sets of tested information by allowing each set to be encoded with distinct contextual cues (Tulving & Watkins, 1974): Proactive interference from the previous sets of information is reduced because of the enhanced discrimination (e.g., Pastotter et al., 2011; Szpunar et al., 2008). Another way to overcome this

cost is by providing a distinct category cue: For example, Unsworth et al. (2013, Experiment 4) showed that providing distinct category cues when two lists of items were to be encoded reduced interference between the two lists by narrowing the mental search set at recall, as indicated by reduced between-list intrusions as well as reduced recall latencies (see also, e.g., Gardiner et al., 1972; Wixted & Rohrer, 1993). By restricting memory access to irrelevant information, a category cue also appears to act as a distinctive contextual cue that enables better discrimination between sets of items and therefore greater efficiency at retrieval (see also Thomas & McDaniel, 2013).

Changes in contextual cues more generally can also invoke release from proactive interference buildup. To follow from the discussion in Chapter 3, in list-method directed forgetting, a cue to forget prior information appears to induce a rapid mental context change (Sahakyan & Kelley, 2002), which has been shown to reduce the mental search set at retrieval as demonstrated by reduced recall latencies for the list studied after the forget cue (Bauml & Kliegl, 2013). Reset-of-encoding (Pastotter et al., 2008, 2012) and change of encoding strategy (Sahakyan & Delaney, 2003) after receiving an instruction to forget have also been proposed as mechanisms promoting the contextual discrimination which underlies the release from buildup of proactive interference in directed forgetting. Apart from internal context changes, external changes such as studying information in a different physical environment also appear to enhance memory of new information by reducing proactive interference (Dallett & Wilcox, 1968; see also Smith, 1979, 1982; Smith et al., 1978).

To reiterate, memory for both the aloud items and the silent list items is expected to suffer at retrieval due to noisy context reconstruction caused by the existence of the two lists. However, if the benefit of the distinctive encoding of the aloud items stems from an effect of

contextual shift when studied in List 2, then the aloud items are expected to benefit from a release of proactive interference buildup. This would in turn enhance discriminability between the aloud and silent lists, resulting in better memory for the items from the aloud list and a relative cost for the items from the silent list. This is the second way in which context is expected to influence the production effect. Although it may be argued that there should also be a benefit of distinctive contextual encoding for aloud items even when studied in List 1, the reconstruction of context at retrieval in this case is expected to be less consistent because the encoding occurred first (see e.g., Smith, 1979, who showed that participants must be explicitly prompted to use context information at retrieval). Thus, context is expected to have a much smaller effect for enhancing memory for aloud items studied in List 1 relative to when aloud items are studied in List 2.

To provide further evidence in support of the contribution of distinctive contextual cueing to memory for items studied aloud, the present experiment also included four additional trials in which participants studied two lists either both aloud or both silently and were then tested on either List 1 or List 2 (see rows 5-8 in Table 4). On the aloud-aloud trials, production is expected to act as a distinctive contextual cue for the aloud items studied in List 2 even though List 1 was also studied aloud—production would be predicted to lead to a contextual shift because of the explicit separation of the two lists, regardless of the study mode of List 1. Like the silent-aloud trials, then, memory for the aloud items studied in List 2 are expected to benefit from a release of proactive interference buildup from List 1, because of the greater set discrimination that would lead to better reconstruction of context for List 2 at retrieval; a relative memory cost is in turn expected for the aloud item studied in List 1. No difference in memory is expected between the

two lists on the two silent-silent trials because the noisy reconstruction of context should affect the recall of both sets equally.

In summary, in the case of recall testing, the goal of the present study was to redefine the conceptualization of production as an “unusual” encoding process to instead be an effect of distinctive encoding by contextual cueing. A modified mixed-list design was used which allowed the removal of possible task-switching effects from those of distinctive encoding and item-order processing, by studying aloud and silent items in separation. Trials in which two aloud lists or two silent lists were studied were included also. Although both lists are expected to suffer from interference due to noisy reconstruction of context in the dual-list trials relative to the single-list trials, the key prediction is that memory for the aloud items, when studied second, would benefit from a release from proactive interference buildup due to the influence of distinctive contextual cueing afforded by production during encoding.

Method

Participants. Participants were 63 undergraduate students (19 men, 41 women; age range: 17-32 years, mean age: 20.0 years, $SD = 2.4$ years; 2 participants declined to provide their age, 3 participants declined to provide demographic information) from the University of Waterloo, recruited via the Department of Psychology’s research participation system. Ethics approval was obtained from the University of Waterloo Research Ethics Board, and informed consent was obtained from all participants. Participants received course credit in exchange for their participation. Seven participants were tested but excluded from analyses due to one of several factors: clear lack of effort as demonstrated by not recalling any of the studied items on more than half of the experimental trials (3 participants), having a total number of intrusions 2.5

standard deviations greater than the mean (2 participants), or difficulty in understanding instructions (2 participants).¹⁶

Apparatus. The experiment was conducted online at Pavlovia.org running a program built in PsychoPy Version 2020.1.3 (Open Science Tools, Nottingham, England). Participants completed the experiment on their own computers. A researcher monitored participants for the duration of the study via online video using Cisco Webex (Cisco Systems, San Jose, CA) to ensure understanding and compliance with all of the instructions.

Materials. The study materials consisted of the set of 186 words listed in Appendix C. The words were presented in uppercase, black text, in the Arial font, with a size of 0.05 height units (i.e., a height of 5% of the screen height), against a white background. Six words were used for the practice trial, divided into one list of three aloud items and one list of three silent items. The remaining 180 words were divided into 18 lists of 10 words each, to be used in the 10 real experimental trials according to the arrangements depicted in Table 4. Two trials each contained single lists of 10 words, one studied all aloud and one studied all silently: These constituted the baselines. A further four trials each contained two lists of words: On two of these trials, both lists were studied aloud; on the other two, both lists were studied silently. The remaining four trials each contained one list studied aloud, and one list studied silently: On two of these trials, the

¹⁶ A post-hoc power analysis was performed using G*Power Version 3.1.9.4 (Faul et al., 2007) to confirm that if studying List 1 aloud also enhances memory for aloud items relative to List 2 (regardless of its study mode), that a sufficient number of participants had been included to detect this effect. Such a finding would, of course, stand in opposition to the main hypothesis of this study—that only studying List 2 aloud enhances memory relative to List 1 because the distinct “aloud” context is maintained through retrieval—and would suggest that effects other than context are influencing the “unusual” processing of aloud items in the mixed-list production effect observed in recall. The effect size used in this calculation was based on that from the 2 (aloud/silent) × 2 (List 1/List 2 recall) within-subject ANOVA used to analyze the data when they were collapsed based on whether List 2 was aloud or silent ($\eta_p^2 = 0.154$). The reasoning was that if the main hypothesis was incorrect, a similar memory benefit for aloud information might be observed between the two analyses. Based on alpha = .05 and power = 0.95, the appropriate sample size needed to satisfy these parameters was estimated to be N = 25. Therefore, the sample size of N = 56 is more than adequate for investigating the main hypothesis in the current experiment.

aloud list was studied first; on the other two, the silent list was studied first. Assignment of words to lists was randomized anew for each participant, and the order of word presentation in each list was randomized anew for each participant.

Procedure. Participants began by meeting a researcher on the Webex platform and were then provided with the study's web link. The entire set of instructions was given at the start of the experiment. Just prior to each trial, participants were shown a screen showing "Ready??" presented in the same formatting as the to-be-studied words but in lowercase, and at 0.1 height units above the center of the screen; they pressed the spacebar to proceed. This was to ensure that they did not miss the instruction for whether to study List 1 aloud or silently as no breaks were provided between trials. Participants were instructed that this screen was not meant to be a break and that they should proceed immediately.

Each trial consisted of a study phase, a filler task, and a test phase. For the study phase, participants were told that they would be shown a list of words with an instruction to read the words either aloud ("List 1 – Aloud") or silently ("List 1 – Silent") just prior to seeing the words; the instruction appeared on the centre of the screen for 3 s. Silent reading was to be done without moving lips. Participants were told that they should study the words in the indicated manner and that they were to remember them for a later test. They were also told that on some trials they would study a second list of words ("List 2 – Aloud" or "List 2 – Silent" would be shown on the screen). When this happened, they should follow the instruction just as they did for List 1; moreover, they were to remember both lists for the later test. Each word was presented individually for 1 s at the center of the screen and a blank period of 1 s intervened between successive words.

The filler task began immediately after the study phase and lasted for 16 s. Participants were told that a series of three-digit numbers would be presented, and that they were to order the three digits of each number from largest to smallest, typing their answer on their keyboard. The numbers appeared at 0.05 height units above the center of the screen, and participants' typed responses appeared at 0.1 height units below the center of the screen in the same formatting as the numbers. Participants pressed the "Enter" key after they finished each response; their typed response then disappeared from the screen, and the next number was shown. The list of numbers was randomized anew for each participant.

The test phase began immediately after the filler task and lasted for 60 s. Participants were asked to recall the only list that was studied on the single-list trials, or only one of the lists studied on the two-list trials. The instruction "Recall the words from List 1" or "Recall the words from List 2" appeared in the same formatting as the "Ready??" instruction at the start of each trial, at 0.2 units above the center of the screen, and three question marks ("???",) with the same formatting also appeared at 0.05 units above the center of the screen. Participants typed each word that they were able to remember, and their response appeared on the screen as they typed at 0.1 height units below the center of the screen in the same formatting as the other text. They pressed the "Enter" key after they finished typing each word, after which their typed response disappeared from the screen and they typed the next word. The next trial began immediately after the preceding test phase.

The experiment began with a practice trial to familiarize participants with the procedure, using a shorter study phase with two 3-item lists as described in the Materials section. Participants were randomly assigned to recall either List 1 or List 2 in the practice trial. After a final chance to ask any questions, the 10 experimental trials took place. For the four pairs of two-

trials (as described in the Materials section), participants were assigned at random to recall List 1 for one trial and List 2 for the other trial. The order of the 10 trials and designation of the list to be recalled were randomized anew for each participant. The entire experiment took under 1 hour to complete.

Results

The results are shown in Table 5 and Figure 5. The dependent measures are correct recall and number of intrusions; the number of intrusions indicated in Table 5 includes both within- and between-trial intrusions of previously studied items recalled from the wrong list, separated according to whether intrusions were from a list studied aloud or silently. Primary analyses were conducted using JASP Version 0.14 (JASP team, 2020).

Comparing Single Lists to Double Lists

First, the two trials which contained only one list were compared to the two-list trials. Recall performance did not differ between these two single-list trials, $t(111) = 0.43, p = .671, d = 0.057$, nor did the inter-list intrusions, as shown by a non-significant 2 (aloud/silent intrusions) \times 2 (trial) within-subject ANOVA, $F(1, 55) < 0.01, MSE = 0.200, p > .999, \eta_p^2 < 0.001$, with no significant main effect of intrusion type, $F(1, 55) = 1.15, MSE = 0.249, p = .289, \eta_p^2 = 0.020$, nor of trial, $F(1, 55) = 0.10, MSE = 0.181, p = .755, \eta_p^2 = 0.002$. Inspection of Table 5 and Figure 5 shows that recall for these two single-list trials was better than recall of both List 1 and List 2 in the eight two-list trials. This was confirmed by a pair of paired-samples t -tests showing that recall performance in the two single-list trials was higher than recall of both List 1, $t(111) = 8.58, p < .001, d = 0.811$, and List 2, $t(111) = 4.82, p < .001, d = 0.455$, in the two-list trials.

Table 5

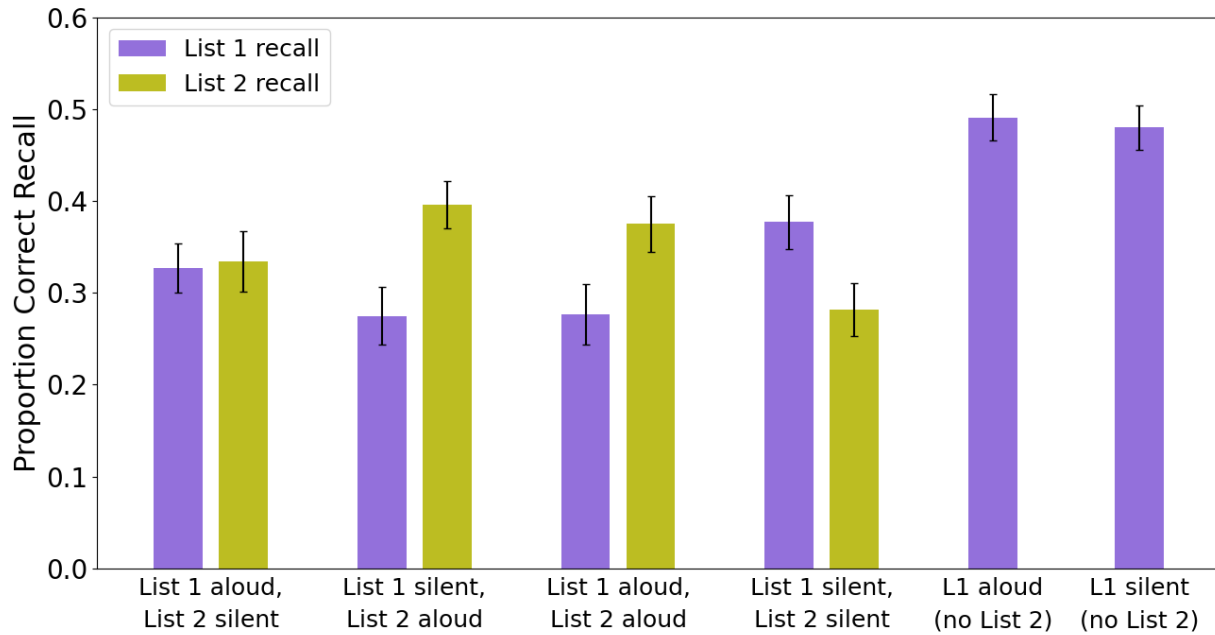
Proportion correct recall and average number of intrusions for each of the 10 experimental trials.

Study order	Proportion of List 1 recall	No. of intrusions		Proportion of List 2 recall	No. of intrusions	
		Aloud	Silent		Aloud	Silent
List 1 aloud, List 2 silent	.327 (.027)	0.13 (0.051)	1.02 (0.215) 0.75 (0.166)	.334 (.033)	1.05 (0.152) 1.05 (0.152)	0.13 (0.051)
List 1 silent, List 2 aloud	.275 (.031)	0.77 (0.155) 0.63 (0.143)	0.14 (0.059)	.396 (.026)	0.20 (0.065)	0.86 (0.152) 0.79 (0.150)
List 1 aloud, List 2 aloud	.277 (.033)	1.29 (0.170) 1.09 (0.160)	0.09 (.0038)	.375 (.030)	0.93 (0.137) 0.82 (0.135)	0.07 (0.035)
List 1 silent, List 2 silent	.377 (.029)	0.05 (0.030)	1.00 (0.169) 0.93 (0.173)	.282 (.029)	0.09 (0.038)	1.11 (0.188) 1.02 (0.176)
List 1 aloud (no List 2)	.491 (.025)	0.20 (0.059)	0.13 (0.063)	-	-	-
List 1 silent (no List 2)	.480 (.024)	0.18 (0.068)	0.11 (0.049)	-	-	-

Note. Standard errors in parentheses. Intrusions are separated by whether the intruding items were studied aloud or silently. Where applicable, intrusions are also separated according to the number of within-trial or between-trial intrusions: The total number of intrusions is indicated above the number of within-trial intrusions.

Figure 5

Proportion correct recall.



Note. Error bars represent standard error.

There were more (total) intrusion errors on the eight two-list trials relative to the two single-list trials both during List 1 recall, $t(111) = 5.02, p < .001, d = 0.474$, and during List 2 recall, $t(111) = 6.45, p < .001, d = 0.610$, although for the two-list trials the number of intrusions between Lists 1 and 2 did not differ, $t(223) = 0.12, p = .907, d = 0.008$. For these two-list trials, most of the intrusions were within-trial: 73% for List 1 and 79% for List 2. Thus, when two lists were studied, there was proactive interference from the study of List 1 onto List 2 and there was retroactive interference from the study of List 2 onto List 1—which appears to be responsible for the reduced memory in these trials compared to the two single-list trials, due to noisy reconstruction of context. Given that inspection of Table 5 reveals differing levels of such interference in terms of recall performance across the eight two-list trials, the next step was to probe the possible differential contributing effects of studying items aloud versus silently on the amounts of interference observed.

Retroactive Interference

Two analyses were performed to examine the overall interference effects from the study of aloud items versus the study of silent items. First, the effect of retroactive interference from List 2 onto List 1 (and thus whether List 2 was protected because of a release of proactive interference buildup from List 1) was compared depending on whether List 2 was aloud or silent. This was done by separately collapsing the two lists where List 2 was aloud and the two lists where List 2 was silent (in Table 5, rows 2 and 3, and rows 1 and 4). In the 2 (List 2: aloud/silent) \times 2 (List 1/List 2 recall) within-subject ANOVA, the main effect of List 2 being aloud versus silent was not significant, $F(1, 111) < 0.01$, $MSE = 0.030$, $p = .956$, $\eta_p^2 < 0.001$, nor was the main effect of recall by list, $F(1, 111) = 3.54$, $MSE = 0.035$, $p = .063$, $\eta_p^2 = 0.031$.

The interaction, however, was significant, $F(1, 111) = 20.25$, $MSE = 0.033$, $p < .001$, $\eta_p^2 = 0.154$, indicating that List 1 and List 2 recall differed as a function of whether List 2 was studied aloud or silently. Given the significant interaction, the simple main effects were examined next. Paired-samples t -tests showed that when List 2 was aloud, recall of List 1 was significantly poorer than recall of List 2, $t(111) = 4.60$, $p < .001$, $d = 0.435$; however, when List 2 was silent, List 1 and List 2 recall did not differ, $t(111) = 1.74$, $p = .084$, $d = 0.165$. In addition, List 1 recall was poorer when List 2 was aloud relative to when List 2 was silent, $t(111) = 3.25$, $p = .002$, $d = 0.307$, whereas List 2 recall was greater when List 2 was aloud relative to when it was silent, $t(111) = 3.26$, $p = .001$, $d = 0.308$.

Proactive Interference

Next, the effect of proactive interference from List 1 onto List 2 (and thus in turn, whether List 1 is protected from retroactive interference due to List 2) was compared depending on whether List 1 was aloud versus silent. To do this, the two lists where List 1 was aloud and

the two lists where List 1 was silent (in Table 5, rows 1 and 3, and rows 2 and 4) were separately collapsed. In this 2 (List 1: aloud/silent) \times 2 (List 1/List 2 recall) within-subject ANOVA, neither the main effect of List 1 being aloud versus silent, $F(1, 111) = 0.08$, $MSE = 0.030$, $p = .784$, $\eta_p^2 < 0.001$, nor the main effect of recall by list, $F(1, 111) = 3.54$, $MSE = 0.035$, $p = .063$, $\eta_p^2 = 0.031$, was significant. The interaction also was not significant, $F(1, 111) = 1.13$, $MSE = 0.038$, $p = .290$, $\eta_p^2 = 0.010$.

The pair of analyses in these two sections suggest that only studying items aloud, but not silently, on the second list (i.e., more recently) interfered with memory of the other list of items (i.e., List 1) regardless of the study mode of List 1. In turn, when studied aloud, List 2 also allowed for a release from buildup of proactive interference from List 1. This would have promoted a better reconstruction of context at retrieval due to the enhanced discriminability between List 1 and List 2, which improved memory of List 2 compared to when it was studied silently. Study mode of List 1 did not influence the amount of interference between lists. Similar analyses performed on the results of intrusions did not yield any significant main effects nor interactions, whether examining only within-trial intrusions or including both within-trial and between-trial intrusions.

Specific Recall on Two-List Trials

Study Mode of List 2. The results of the individual two-list trials were then probed more closely (i.e., without collapsing them) as confirmation for the findings above. First, the effect of the study mode of List 2 was examined. For the four trials in which List 2 was studied aloud, a 2 (study order: silent-aloud/aloud-aloud) \times 2 (List 1/List 2 recall) within-subject ANOVA revealed a main effect of recall by list such that recall of List 1 was poorer than that of List 2, $F(1, 55) = 17.34$, $MSE = 0.039$, $p < .001$, $\eta_p^2 = 0.240$, although there was no main effect of study order, $F(1,$

55) = 0.14, $MSE = 0.040$, $p = .714$, $\eta_p^2 = 0.002$, and the interaction was not significant, $F(1, 55) = 0.30$, $MSE = 0.025$, $p = .587$, $\eta_p^2 = 0.005$. For the four trials in which List 2 was studied silently, a 2 (study order: aloud-silent/silent-silent) \times 2 (List 1/List 2 recall) within-subject ANOVA revealed no main effect of recall by list, $F(1, 55) = 3.04$, $MSE = 0.035$, $p = .087$, $\eta_p^2 = 0.052$, no main effect of study order, $F(1, 55) < 0.01$, $MSE = 0.030$, $p = .970$, $\eta_p^2 < 0.001$, and a non-significant interaction (after applying a Bonferroni correction), $F(1, 55) = 4.35$, $MSE = 0.033$, $p = .042$, $\eta_p^2 = 0.073$.

These results confirm that, relative to studying List 2 silently, studying List 2 aloud resulted in retroactive interference that influenced memory for List 1, and thus in turn protected List 2 items from proactive interference from List 1, regardless of the study mode of List 1. Studying List 2 silently resulted in neither differing interference nor protective effects relative to those effects coming from studying List 1, again regardless of the study mode of List 1. Therefore, when List 2 was studied aloud, the distinctiveness aspect afforded by studying items aloud appears to have been especially helpful in invoking a release from buildup of proactive interference from studying List 1, allowing for a better-focused reconstruction of context of List 2 at retrieval and thereby enhancing memory of List 2.

Study Mode of List 1. Next, the examination turned to the four trials in which List 1 was studied aloud, and then the four trials in which List 1 was studied silently. When List 1 was studied aloud, a 2 (study order: aloud-silent/aloud-aloud) \times 2 (List 1/List 2 recall) within-subject ANOVA revealed non-significant main effects of recall by list (after applying a Bonferroni correction), $F(1, 55) = 4.33$, $MSE = 0.036$, $p = .042$, $\eta_p^2 = 0.073$, and of study order, $F(1, 55) = 0.03$, $MSE = 0.035$, $p = .859$, $\eta_p^2 < 0.001$; the interaction was also non-significant, $F(1, 55) = 3.43$, $MSE = 0.034$, $p = .069$, $\eta_p^2 = 0.059$. This shows that when aloud items are studied first,

there are no differing interference and protective effects relative to those effects coming from studying List 2, regardless of the study mode of List 2.

In contrast, when List 1 was studied silently, a 2 (study order: silent-aloud/silent-silent) \times 2 (List 1/List 2 recall) within-subject ANOVA again revealed non-significant main effects of recall by list, $F(1, 55) = 0.28$, $MSE = 0.035$, $p = .596$, $\eta_p^2 = 0.005$, and of study order, $F(1, 55) = 0.07$, $MSE = 0.032$, $p = .796$, $\eta_p^2 = 0.001$, but here the interaction was significant, $F(1, 55) = 23.59$, $MSE = 0.028$, $p < .001$, $\eta_p^2 = 0.300$. Therefore, the relevant simple effects were probed. Paired-samples t -tests showed that for the two silent-silent trials, List 1 recall was significantly higher than List 2 recall, $t(55) = 2.88$, $p = .006$, $d = 0.385$, and List 1 recall was also significantly better than in the silent-aloud trials, $t(55) = 3.16$, $p = .003$, $d = 0.423$, whereas List 2 recall was significantly poorer than that in the silent-aloud trials, $t(55) = 3.44$, $p = .001$, $d = 0.459$.

This marked difference between when both lists were studied silently and when List 1 was silent but List 2 was aloud highlights a possible effect of order that may be particular to studying both lists silently: The encoding of order information from List 1, as a product of silent studying, prevents order information from being encoded in List 2 due to a greater buildup of proactive interference specifically caused by the encoding of order information. As a result, List 1 is also relatively more protected from retroactive interference that would otherwise arise from studying List 2, but only when List 2 is studied silently. A discussion of this finding in more detail is presented next.

Discussion

The goal of this study, in keeping with the main goal of the dissertation, was to examine the production effect in recall as a possible context-based memory effect. Previously, the production effect in recall has been framed as arising due to production being an “unusual”—

distinctive—encoding process relative to silent reading. Here, the aim was to seek a more concrete description of what this distinctive encoding process entails. To that end, it was predicted that the distinctive processing afforded by the produced items represents a global contextual framework that becomes associated with (only) the items produced aloud during encoding, which enables more accurate reconstruction of context at retrieval to improve memory for the produced items. This mechanism is presumed to consequently interfere with memory of the relational information for the unproduced items due to a release from buildup of proactive interference, resulting in a reduction in maintenance because of the contextual mismatch, and thus a relative memory cost for the unproduced items.

A modified mixed-list design was used, which permitted the removal of possible task-switching effects within the typical random mixed-list production effect design that would otherwise be confounded with the processes of distinctive encoding (for the aloud items) and item-order memory (for the silent items). To accomplish this, participants studied aloud and silent items separately in pairs of lists. Trials in which two aloud lists or two silent lists were studied were also included, and the results from all of these dual-list trials were compared with two single-list trials in which only an aloud list or only a silent list was studied. For trials on which a pair of lists was studied, only one list was tested so that proactive and retroactive interference effects could be examined separately.

It was expected that both lists in the dual-list trials would suffer from memory costs relative to the single-list trials due to noisy reconstruction of context at retrieval (Unsworth et al., 2012, 2013). However, it was predicted that recall of the aloud list, when studied as List 2, would benefit from a release from proactive interference buildup due to the distinctive contextual cueing afforded by production during encoding (which, as stated above, would allow a more

accurate reconstruction of context at retrieval relative to List 1). Consequently, this mechanism should overcome some of the expected interference from List 1 in the overall recall of a List 2 that had been studied aloud. In contrast, when List 2 was studied silently, no difference in memory was expected between the two lists. In particular, if List 1 was studied aloud, studying List 2 silently should benefit from an enhanced memory for item order because it is studied more recently, compensating for the benefit of—and the interference from—production accruing to List 1 because the reconstruction of context for List 1 would be less effective (e.g., Smith, 1979).

Interference and Contextual Differentiation between Aloud and Silent Lists

To begin with, performance did not differ between the two single-list trials—one studied entirely aloud and the other studied entirely silently—which is consistent with previous findings showing absence of between-subjects production effects in recall (e.g., Jonker et al., 2014). This finding thus provides evidence in favour of a dual-mechanism encoding account for the production effect found in recall—that distinctive processing supports the encoding of the produced items whereas serial order memory separately supports the encoding of the unproduced items. In addition, performance for both lists within the dual-list trials was worse relative to that of the single-list trials, confirming the prediction that noisy reconstruction of temporal-contextual cues occurs at retrieval when recalling only one set of information learned within multiple sets. The result is interference and reduced recall of all sets of items compared to when only a single set of information needs to be remembered (Unsworth et al., 2012, 2013). In support of this observation, the findings also showed that the majority of incorrectly recalled items—intrusions—in the dual-list trials were within-trial intrusions (i.e., items recalled from the other list within the same trial) rather than between-trial intrusions, indicating that interference primarily came from the non-recalled list studied within the same trial.

Consider next the main prediction regarding memory for the aloud items when studied as one of the two lists in a trial. Due to the distinctive encoding of aloud items when studied in List 2, a new contextual framework is expected to form for these aloud items, leading also to a release from proactive interference buildup from List 1. This in turn is expected to result in a reduction in memory maintenance of the List 1 items because of the contextual change (particularly that of item-order memory if List 1 is studied silently). List 2 should consequently benefit from a more accurate reconstruction of contextual cues at retrieval compared to List 1, improving memory of the List 2 items. Collapsing across trials based on the study mode of List 2 (i.e., without regard to the study mode of List 1), the results showed that when List 2 was studied aloud, memory of List 2 was enhanced relative both to List 1 and to List 2 when it was studied silently. Conversely, memory for List 1 suffered from a memory cost relative to List 2, and to List 1 when List 2 was studied silently. Having studied List 2 aloud thus appeared to create a distinctive context that was still in place at the time of recall, resulting in the observed memory benefit for the aloud items that became associated with that distinct context. In contrast, it became more difficult to reconstruct the relevant temporal-contextual cues from List 1 because the formation of the distinct context for List 2 led to a reduction in memory maintenance of the List 1 items (i.e., because the List 1 items are not associated with this context), leading to a relative memory cost for the List 1 items.

Collapsing across trials based on the study mode of List 1, the analyses showed that there were no memory differences between Lists 1 and 2 based on whether List 1 was studied aloud or silently. These results are consistent with the prediction that the reconstruction of context is less effective for information studied earlier, regardless of study mode—and thus study mode of List 1 had no differential effect on memory for List 2.

In a typical mixed-list production effect design with a recall test, the memory benefit for items studied aloud appears to offset cross-interference effects from the items studied silently, and performance for the aloud items becomes similar to that of studying pure lists while silent items appear to suffer from a memory cost. In the present experiment, however, performance for aloud items studied in List 2 still fell short of that on the single-list trials, perhaps due to the overall greater length of the modified mixed-list design and to the influence of the preceding trials on later trials (i.e., between-trial intrusive effects). The lack of any effects on memory based on the study mode of List 1 is also consistent with the hypothesis that it is the distinctive encoding process of the aloud items that disrupts the (item-order) learning mechanisms of items studied silently, and thus production effects in recall should disappear when this disruptive influence is removed (i.e., when silent items are studied after the aloud items).

Interestingly, while memory was relatively enhanced for the items studied aloud in List 2, this did not lead to a corresponding pattern of results concerning the number of intrusions during recall— if List 2 was aloud, one would perhaps expect more intrusions in List 1 than in List 2 recall, as well as more intrusions than if List 2 was silent (though the latter was certainly observed numerically: The average number of within-trial intrusions was 0.81 when List 2 was aloud, versus 1.04 when List 2 was silent). Given that in the present data the numbers of intrusions were primarily between 0 to 2 per trial, it may be that there were simply not enough intrusions to reveal any particular patterns—future studies will require a greater number of observed intrusions for meaningful investigation.

Generalizability to Mixed-List Designs

As briefly discussed in Chapter 1, the present findings can be generalized to the case of encoding mixed lists containing both aloud and silent items. The release from proactive

interference buildup for the aloud items, occurring in the present study only when they are studied in List 2, is believed to be the key mechanism for the formation of production as a distinct contextual framework even for aloud items studied in mixed lists. In these mixed-list designs, this release is hypothesized to occur *each* time an item is studied aloud. This encoding mechanism is presumed to be identical in mixed-list designs examining recognition memory, since the *encoding* phase remains identical regardless of test method. But what happens to aloud items studied earlier if a later aloud item causes a “release” from earlier items within the same list? The hypothesis is that in mixed-list designs, only silent items would be “released” through the study of later aloud items—the aloud items that are already studied would be maintained because they are all studied within one “list” and associated under a common contextual framework of production.

This is the key distinction between mixed-list designs and the aloud-aloud trials (Table 5, row 3) in the current experiment—the aloud lists in these trials were considered separate lists, and hence the aloud items from List 1 were subject to release due to the encoding of List 2. To further generalize to other encoding methods apart from production, in any mixed-list design where multiple encoding methods are used, the hypothesis is that the most distinct encoding method will form the primary contextual framework in a similar manner—for example, generation appears to “prevail” over production when used together (Nairne et al., 1991). This hypothesis remains speculative and requires future confirmation.

As the present study also aims to show, the retrieval method would influence how the effect of production on memory is revealed. In recognition tests but not in recall tests, the studied information is presented again at retrieval. On a recall test, then, the release from proactive interference buildup due to the act of production may be especially likely to result in an apparent

memory cost for the information studied silently, when studied within the same list as aloud items. The silent items are continuously dropped from memory maintenance because of the production-induced contextual change whenever an aloud item is studied, and also must be recalled without seeing the items again at retrieval—a process that coincides with the selective impact on memory for item order within the silent items (which is irrelevant in recognition tests, where the experimenter controls item sequence). This stands in contrast with the observation that in recall, memory for items studied silently in pure lists is typically similar to that of aloud items, because item order can be an effective technique for retrieving silently studied information, as discussed above. Relatedly, the irrelevance of order information may contribute to the worse memory for silent information relative to aloud information when these are tested via recognition.

Aloud items would also suffer from a relatively smaller memory cost due to interference from the concurrent study of silent items. However, the formation of production as a distinct contextual framework appears to offset this interference, improving memory of the aloud items from a mixed list up to par with that when encoding pure aloud lists—but only when testing for recall. It would appear that production is used more consistently as a contextual cue when recall, rather than recognition, is required of pure aloud lists: In recall, memory for aloud items is typically similar whether studied in mixed or pure lists; in contrast, in recognition, memory for aloud items has typically been lower when studied in pure lists compared to mixed lists (e.g., MacLeod et al., 2010). The results in Chapter 2 of course showed that the latter appears to be due to the inconsistent use of production as a contextual cue when memory for pure aloud lists is tested via recognition.

The Case of Studying Two Silent Lists

Compared to the analyses of the collapsed trials, the more detailed trial-by-trial analyses revealed a largely consistent pattern of findings—namely, memory for aloud items studied in List 2 was enhanced without regard for whether List 1 was studied aloud or silently.

Surprisingly, however, one result stood in contrast to the previous analyses as well as to the initial predictions: When two lists were both studied silently within one trial, memory for List 1 was better than List 2. This result is inconsistent with those reported in Unsworth et al. (2013) and suggests that additional mechanisms in the present experiment contributed to memory for the silent items apart from the influence of temporal-contextual cueing.

To account for the benefit for silent items when studied in List 1, a hypothesized mechanism aimed at describing the relation between the encoding of item versus order information proposed by Guitard et al. (2021) was considered. In their study, participants were given two lists of items to learn on each trial and were instructed to remember for each list either the items themselves or their serial order information. Participants were then tested on one of the lists (each participant received all combinations of item and order learning in a within-subject design). The findings showed that when different modes of encoding were required for the two lists (e.g., item memory for one list and order memory for the other list), memory for serial order was enhanced—particularly on List 2—relative to the trials in which encoding of serial order was required for both lists. Manipulations of encoding modes did not result in similar, consistent differences when item memory was tested. This suggested that item memory and order memory are maintained via at least some distinct processes in working memory (see Majerus, 2019, for a review), rather than sharing entirely common maintenance processes (see also, e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998; Nairne, 1988, 1990, for modelling that suggests

that there are separate representations of item and order information within short-term memory). Critically, working memory capacity for retaining serial order is believed to be more limited than its capacity for retaining item information. Put another way, maintaining order information appears to impose a heavier load on working memory than does maintaining item information.

It is hypothesized that, in the present study, when two lists within one trial were both studied silently, the items in List 1 may have been better remembered due to the encoding of order information for List 1 resulting in a subsequent reduction of working memory capacity for the encoding of order information for List 2. However, this is inconsistent with the results in Guitard et al. (2021) showing that when participants were required to remember the serial order for both lists within a trial, memory for List 2 was overall better than memory for List 1 (except in their Experiment 3 where there was no difference in memory). These results are also related to the prediction made in Guitard et al. (2021) that the memory cost for List 1 was due to the earlier-encoded order information from List 1 being “overwritten” by the encoding of order information in the List 2 items—essentially, a form of retroactive interference.

Two main differences between the current experimental procedure and those in Guitard et al. (2021) may have led to the inconsistent findings. First, because participants in the present study were aware that they would be required to study some information aloud—which was more easily remembered—they may have shifted their focus on the silent-silent trials to try to aid their memory for List 1 in case List 2 had to be studied aloud. In addition, the present study used free recall tests whereas Guitard et al. (2021) primarily tested for recognition memory (Experiments 1-4) or, in one experiment, cued recall (Experiment 5): As noted in the earlier chapters, the influence of production appears to differ on recall versus recognition tests. Future studies should continue to investigate the possible contributions of distinctive encoding

processes such as production, as well as different methods of retrieval, to how people may try to focus their encoding of order information for multiple sets of information.

In sum, the findings in this chapter provide further evidence in support of the prediction that production can act as a distinct contextual cue to aid recall of information studied aloud. In a mixed list production effect design, while both the aloud and the silent items interfere with each other during encoding, the distinctive contextual framework formed by production improves the retrieval process of items studied aloud relative to items studied silently. This occurs because of a release from proactive interference buildup, resulting in reduced maintenance of item-order memory and a consequent overall memory cost for the silent items because of the contextual change process, while the aloud items benefit from a more accurate reconstruction of context at retrieval. Moreover, the results support the view that there appears to be a more limited capacity within working memory for maintaining serial order information compared to item information—although the exact process for how a section of order information is chosen to be maintained remains to be investigated.

Chapter 5 – General Discussion

Across three converging manipulations, this dissertation examined whether the production effect—the memory benefit from producing information during encoding (typically by reading aloud)—could be characterized more broadly as a context-based memory effect. The investigation focused on how the distinctiveness of produced information may lead to the formation of an associated global contextual framework, which contributes toward production effects observed in both recognition and recall by aiding memory for produced information via contextual cueing.

In brief, the first study (Chapter 2) investigated whether, by implicitly showing participants that the distinctiveness heuristic can be useful as a context-based memory technique even when pure lists containing only aloud items or only silent (i.e., unproduced) items are studied, a robust production effect can be elicited in the study of pure lists, close to that seen when studying mixed lists. The rationale was that if the production effect is context-based, production *by itself* should improve memory for items studied aloud regardless of whether contrasting silent items are studied concurrently: The key factor for the memory boost is the consistent use of production as a contextual cue in any given situation. A secondary finding suggested that the proposed contribution of statistical distinctiveness to the production effect may be better explained as an effect of contextual cue overload, as it was shown that only a change in the absolute, but not relative, number of produced items influenced memory of these aloud items and therefore the magnitude of the production effect. Rather than being attributed to the relative frequency of aloud versus silent items, an influence of statistical distinctiveness in the production effect may occur because of changes in the specificity of production as a contextual cue when the actual number of aloud items studied increases or decreases.

The second study (Chapter 3) tested for evidence in support of the context hypothesis using the contextual change account of directed forgetting, supposing that instructing participants to remember intentionally forgotten items based on whether they were studied aloud or silently reinstates the use of production as a critical contextual cue established at encoding, which should improve memory of items studied aloud and attenuate effects of directed forgetting. Whereas the first two studies examined the production effect as measured by recognition tests, the third study (Chapter 4) examined how the distinctiveness of aloud items, as a context-based effect, interferes with the study of other items when the test is recall. The hypothesis was that when the items studied most recently prior to testing were studied aloud, the formation of production as a distinct context associated with these items would reduce memory of information studied previously regardless of study mode, occurring through a release of proactive interference buildup from the information studied earlier.

Summary of the Findings

Studies have shown that production effects can be observed on recognition tests using between-subjects, pure-list designs, where each participant studies only produced or only unproduced information (Fawcett, 2013). However, in this design, the production effect typically is smaller relative to the effects observed using within-subject, mixed-list designs, where each participant studies both produced and unproduced items (see MacLeod & Bodner, 2017, for a brief review). The goal of the first study (Chapter 2) was to examine the prediction that the smaller production effects observed under pure-list designs may be due to an inconsistent application of the distinctiveness heuristic (Dodson & Schacter, 2001): The memory technique wherein remembering that an item was studied aloud facilitates memory for the item itself. The distractor items on a recognition test can be differentiated from the items studied aloud by virtue

of not having been produced, so ensuring that participants are made aware of the usefulness of the distinctiveness heuristic when studying pure lists was predicted to increase the overall magnitude of the production effect in pure lists.

In the first study, additional pure-list items—read either all aloud or all silently—were alternated with items processed either aloud versus silently, the latter manipulated within-subject in a mixed-list procedure. The hypothesis was that through the co-occurring mixed-list manipulation, participants would be made aware (consciously or unconsciously) that they could use the distinctiveness heuristic to identify the pure-list aloud items on a recognition test, thereby improving their memory for these items. Participants were instructed that the mixed-list items would be tested but were falsely led to believe that the pure-list items would not be tested, which acted to differentiate the two types of lists. When the pure-list items were in fact tested, a quite large between-subjects production effect was observed, supporting the main hypothesis: Production of some of the mixed-list items invoked the use of production as a contextual cue that enhanced memory of the pure-list aloud items along with the mixed-list aloud items.

Additionally, a secondary result suggested that because reading all of the pure-list words aloud effectively increased the overall proportion of aloud words, the statistical distinctiveness of the to-be-remembered aloud words in the mixed list was thereby decreased. Correspondingly, in this condition, there was a decrease in the magnitude of the production effect in the mixed-list items relative to a baseline with no pure-list items, due to reduced memory for the aloud words. When the pure-list words were all read silently (thus decreasing the overall proportion of aloud words), however, the magnitude of the production effect was unchanged relative to the baseline: Memory did not differ for either the aloud or silent items. These results provided partial support for the influence of statistical distinctiveness on the magnitude of the production effect based on

the relative frequency of the aloud versus silent items (Icht et al., 2014)—the original intention of Chapter 2.

I hypothesized post-hoc that there was better support for a context-based explanation for these findings due to the consistency of these results with the predictions of the cue overload hypothesis in the existing context literature (e.g., Smith & Manzano, 2010): Greater numbers of items associated with a particular context result in reduced specificity of the contextual association, thereby decreasing the effectiveness of contextual cueing during retrieval. As reported, only an increase in the *absolute* number of aloud items studied relative to the baseline (i.e., when the pure-list items were studied aloud) affected the magnitude of the production effect by decreasing overall memory for the aloud items. A decrease in the *relative* number of aloud items studied (i.e., when the pure-list items were studied silently)—in which the number of aloud items remained the same as the baseline—did not influence memory of either the aloud or silent items. The suggestion is that the apparent partial influence of statistical distinctiveness when the pure-list items were studied aloud was due to the production contextual framework being more overloaded in this experimental condition relative to the baseline.

The view that production serves as a contextual framework within recognition memory was tested more directly in the second study (Chapter 3), which employed the list method procedure from directed forgetting. Two mixed lists, each containing both aloud and silent items, were studied. Half of the participants were falsely led to believe that List 1 would not be tested so it could be forgotten, the typical procedure in list-method directed forgetting. For the aloud items specifically, the contextual change account of directed forgetting (Sahakyan & Kelley, 2002) predicts that the effect of the instruction to forget should be attenuated when participants are instructed to remember the words based on how they studied them (the Aloud versus Silent

Differentiation group). This test instruction should reinstate the distinct contextual information—production (i.e., the source information for List 1 aloud items in particular)—that is theorized to function as a memory aid for items studied aloud within the production effect paradigm. In contrast, instructing participants at test to remember the words based on the list that they had appeared in (the List 1 versus List 2 Differentiation group) does not reinstate the relevant distinct contextual information and therefore should lead to typical directed forgetting effects on the recognition test (cf. Lehman & Malmberg, 2009).

In the List 1 versus List 2 Differentiation group, the hit rate for List 1 was lower in the forget group than in the remember group, whereas the hit rate for List 2 was higher in the forget group than in the remember group—the typical directed forgetting cost and benefit effects seen in that literature. In sharp contrast, no such interaction was found for the Aloud versus Silent Differentiation group. This different pattern supports a contextual characterization of the production effect: Because people do establish source memory for studying aloud items, when the test is recognition, production can then be used as associative contextual information to overcome the effect of directed forgetting. The silent items showed an overall benefit for List 2 recognition, which offers support for a dual-mechanism perspective of directed forgetting in which List 2 enhancement occurs because of additional encoding processes, such as reset-of-encoding (Pastotter et al., 2012).

Previous production effect studies using recall tests have suggested that when people study a mixed list, the distinctiveness—or “uniqueness”—aspect of studying items aloud disrupts list order information during the encoding of silent items, leading to a memory cost for these silent items (e.g., Jonker et al., 2014). To provide more clarity for this mechanism, the third study (Chapter 4) investigated retroactive versus proactive interference effects caused by

studying aloud versus silent items. Specifically, the question was whether interference from the distinctiveness due to studying some items aloud constitutes a possible effect of contextual formation. Participants studied (1) four trials each consisting of one aloud list and one silent list (in two trials, the aloud list was studied first; in the other two trials, the silent list was studied first), (2) two trials each consisting of two aloud lists, and (3) two trials each consisting of two silent lists. There were also two control trials with only one aloud or one silent list. In the trials where two lists were studied, testing of List 1 or List 2 was alternated across trials such that each list was tested once but the order was random.

In all of the two-list trials, there was proactive interference from List 1 onto List 2, as well as retroactive interference from List 2 onto List 1, the typical pattern (cf. Unsworth et al., 2013). Relative to studying List 2 silently, however, studying List 2 aloud invoked greater retroactive interference onto List 1 items, regardless of the study mode of List 1. In contrast, this encoding effect difference was not found when examining the possible proactive interference effects of study mode of List 1. When a list of aloud items is studied second, the hypothesis was that the greater distinctiveness of these aloud items produces a release from buildup of proactive interference from previously studied items because a change of context occurs when List 2 is studied aloud (regardless of the study mode of List 1). This release mechanism thus improves memory for these aloud items. It was additionally hypothesized that the study of mixed-lists under within-subject manipulations could invoke a similar mechanism. Interestingly, only when two silent lists were studied did List 1 show greater proactive interference onto List 2. This suggested that, in certain instances, the encoding of list order information results in a buildup of proactive interference and protects earlier items from future interference effects, but only when additional items are also studied silently.

The next section sets out, in more detail, the contributions of these findings to the theoretical perspective that the production effect can be characterized as a context-based memory effect.

Theoretical Perspectives: Production and Context

The three studies presented in this dissertation converge to support the central hypothesis that the production effect can be more broadly characterized as a context-based memory effect. Under this hypothesis, production, as a specific encoding method, takes on the role of a global contextual mental framework, facilitating memory for information studied aloud as an associative contextual cue both when the method of retrieval is recognition and when it is recall. In this section, I first summarize the evidence that supports this theoretical perspective when the test is recognition and then consider situations in which the test is recall.

When encoded information is tested using recognition, the distinctiveness account of the production effect proposes that remembering that information was produced, as a form of source memory, facilitates retrieval of that information by creating an additional pathway for retrieval, a pathway that confirms initial study (e.g., MacLeod et al., 2010). Whether this distinctiveness heuristic can be characterized as a context-based memory effect was examined in two situations. First, a context-based technique should be useful not only in within-subject, mixed-list designs (the procedure used most often in the production effect literature), but also in between-subjects, pure-list designs. Ensuring the consistent use of production as a contextual cue in pure-list designs, thereby improving memory for items studied aloud, was proposed to be the basis for increasing the magnitude of the production effect observed in these situations. Second, as a distinctive contextual cue, prompting people to remember whether items were studied aloud at

retrieval should be more effective in facilitating memory for produced items compared to using another non-distinct contextual cue.

The results from Chapter 2 support the first hypothesis. Through the concurrent encoding of mixed-list items, participants appear to have discovered that production can be associated even with the “pure-list” aloud items as mental scaffolding to facilitate later retrieval of these items. In this case, the non-studied distractor items on the recognition test—which were, of course, not produced—were differentiated by means of not being associated with the “production” contextual framework. The result was that a much larger pure-list production effect was observed in this study compared to the previous literature (see Fawcett, 2013, for a review). These findings suggest that the distinctiveness heuristic presents itself as a standalone context-based memory technique for remembering produced information, without having to rely on the presence of different encoding processes (such as reading silently) to elicit the desired effect. In mixed-list designs, the main function of studying some items silently thus appears to be making sure that people are aware of this memory technique. That is, the greater production effects typically observed in mixed-list relative to pure-list designs may be attributed to participants using the distinctiveness heuristic more consistently in the mixed-list case: Having studied aloud and silent items concurrently would be more likely to make people aware that attempting to remember whether items were produced would be useful for retrieving the associated produced information. Whether in the study of pure lists or mixed lists, the key mechanism of production is that it represents a global contextual framework that enables better memory for produced items through associative contextual cueing.

In addition, the secondary results concerning the influence of statistical distinctiveness on the production effect were hypothesized to actually provide greater support for a context-based

explanation of these findings, due to the consistency of these results with the predictions of the cue overload hypothesis in the existing literature examining the effects of context on memory (e.g., Earhard, 1967; Rutherford, 2004; Smith & Manzano, 2010; Watkins & Watkins 1975). This cue overload hypothesis states that with an increasing number of items that become associated with a particular context, the specificity of the contextual association is reduced as the search set increases, thereby decreasing the effectiveness of contextual cueing during retrieval.

In Chapter 2, increasing the *absolute* number of aloud items encoded relative to the baseline (i.e., when the pure-list items were studied aloud) reduced the magnitude of the production effect by decreasing overall memory for the aloud items. In contrast, decreasing the *relative* number of aloud items encoded (i.e., when the pure-list items were studied silently)—while the actual number of aloud items remained the same as the baseline—did not affect the magnitude of the production effect: Memory of both the aloud and silent items remained unchanged. Considered together, these findings were inconsistent with the hypothesis for the contribution of statistical distinctiveness presented in Icht et al. (2014), who posited that the magnitude of the production effect should vary according to changes in the *relative* frequency of the aloud versus silent items within a study set: No change occurred in the magnitude of the production effect when the relative number of silent items was increased.

The apparent partial influence of statistical distinctiveness when the pure-list items were studied aloud can be plausibly attributed to the production contextual framework being more overloaded relative to the baseline in this condition: A greater number of aloud items were associated with the context of production. There was certainly no such influence of overloading when the additional pure-list items were studied silently, and hence no change in the magnitude of the production effect in this condition. Production thus possesses another characteristic

associated with context-based memory effects—cue overloading—which provides further support for the central hypothesis that the production effect belongs to this domain of memory effects. On this line of reasoning, cue overloading effects when there is a change only in the total number of encoded items, but not in the number of items associated with a particular context (as was the case when the additional pure-list items were studied silently), remain to be investigated further.

The findings from Chapter 3 support the second hypothesis. Here, the list method procedure from directed forgetting was used to test whether reinstating the use of production as a distinct contextual cue at retrieval would enhance memory for aloud information intentionally forgotten, as predicted by the contextual change account of directed forgetting (Sahakyan & Kelley, 2002). Requiring participants to differentiate items at retrieval based on whether they had been studied aloud or silently (the Aloud versus Silent Differentiation condition) attenuated the expected effects from directed forgetting for the aloud items. In contrast, when participants differentiated the studied items based on the list in which they were studied (the List 1 versus List 2 Differentiation condition), the typically expected effects from directed forgetting persisted. These findings suggest that production indeed represents a distinctive contextual framework with which produced items become associated, which then facilitates memory of the intentionally forgotten produced items when participants are prompted to use it at retrieval. Otherwise, these produced items would suffer from forgetting because of the occurrence of a mental contextual change from being instructed to forget List 1 prior to studying List 2. This was exactly the case when participants were prompted to differentiate the studied items by list rather than by whether they were studied aloud versus silently: List information was not a distinct contextual cue for improving memory of the studied information, and hence the relevant contextual information

that would facilitate memory of produced items was not reinstated (cf. Lehman & Malmberg, 2009). The use of the distinctiveness heuristic allowed participants instructed to forget List 1 to essentially join the two studied lists into one list to be remembered because of the common contextual framework associated with the produced items.

The production effect observed in recognition memory can thus be conceptualized as a context-based memory effect and the distinctiveness heuristic as a technique that uses an associative contextual cue to facilitate memory of produced information. In Chapter 2, the results showed that production can enhance memory of produced items without the need of contrasting encoding processes (such as reading silently): Production is the key contextual cue that one needs to use. An apparent cue overloading effect for production when an increasing number of aloud items are studied also supports a context-based explanation of the findings. In Chapter 3, whereas invoking the distinctiveness heuristic in the Aloud versus Silent Differentiation condition served as an associative contextual cue that facilitated memory of the intentionally forgotten List 1 aloud items, in the List 1 versus List 2 Differentiation condition, List 1 context was dissimilar to that of List 2 and of the retrieval phase, resulting in forgetting of the List 1 items but a memory enhancement for the List 2 items—the expected pattern of directed forgetting.

Chapter 4 then explored the role of production as a context-based memory effect when the test was recall. The central question was whether, in recall, production effects stem from the formation of a distinctive contextual setting across encoding and retrieval: Production would act as the distinct contextual cue that facilitates memory of produced items but would interfere with the encoding and retrieval of silent items. Recall that participants study aloud and silent items in separate lists. Because of the distinctive encoding of aloud items when they were studied second

(in List 2), a new contextual framework was formed for these aloud items, leading also to a release from proactive interference buildup from the earlier studied items (List 1). This resulted in a reduction in memory maintenance of the List 1 items because of the contextual change—particularly that of item-order memory if List 1 is studied silently. List 2 consequently benefitted from a more accurate reconstruction of contextual cues at retrieval compared to List 1, improving memory of the List 2 items. When List 2 was aloud, it thus conferred a distinctive context that was still in place at the time of recall, resulting in the observed memory benefit for the aloud items that became associated with that distinct context. At the same time, it became more difficult to reconstruct the relevant temporal-contextual cues from List 1 because the formation of the distinct context for List 2 led to a reduction in memory maintenance of the List 1 items (i.e., because the List 1 items were not associated with this context), leading to a relative memory cost for the List 1 items. In support of this hypothesis, there was no production effect when silent items were studied after aloud items: The distinctive encoding process of the aloud items that disrupts the (item-order) learning mechanisms of silent items was no longer in place.

Although in this experiment the aloud and silent items were studied in two separate lists, the hypothesis, as discussed in Chapter 4, is that the present findings can be generalized to the case of encoding mixed lists containing both aloud and silent items. The release from proactive interference buildup for the aloud items, occurring in the present study only when they are studied in List 2, is hypothesized to occur *each* time an item is studied aloud, even within mixed-list designs. Only silent items would be “released” through the study of later aloud items—the aloud items that are already studied would be maintained because they are all studied within one “list” and associated under a common contextual framework of production. This context would then be “carried over” to the retrieval phase, facilitating memory for aloud items. This encoding

process is presumed to be identical regardless of whether the test is recall or recognition, given that the encoding phase remains identical in both.

The retrieval method would then influence how the effect of production on memory is revealed. In recognition tests, the studied information is presented again at retrieval—not so in recall tests. On a recall test, it would be especially likely to observe an apparent memory cost for silent items because of the production-induced proactive interference buildup releases when aloud items are studied concurrently. The silent items are continuously dropped from memory maintenance whenever an aloud item is studied because of the production-induced contextual change, and also must be retrieved without seeing the items again at the time of test. This process coincides with the selective influence on memory for serial order within the silent items (which is irrelevant in recognition tests). This stands in contrast with the typical observation wherein, in pure lists, recall for items studied silently is similar to that of items studied aloud because item order can be an effective technique for retrieving silently studied information when it is not interrupted by aloud items (e.g., Jonker et al., 2014). Aloud items would also suffer from a relatively smaller memory cost due to interference from the concurrent study of silent items, although the formation of production as a distinct contextual framework appears to offset this interference: Memory for aloud items in mixed lists is typically up to par with that when encoding pure aloud lists. This would seem to indicate that production is used more consistently as a contextual cue when recall, rather than recognition, is required of pure aloud lists, an idea that makes sense given the reduced environmental support in recall relative to recognition.

Limitations and Future Directions

Limitations and related future aims for each of the three studies were discussed in detail in the individual chapters. Consequently, I will provide only a brief summary of the main

limitations here. In Chapter 2, a main concern was whether the pure-list aloud and silent items were sufficiently differentiated from the concurrently studied mixed-list items, and not routinely combined as one set of items during encoding. The design was not straightforward given that the goal was not to explicitly instruct participants to use the distinctiveness heuristic for the retrieval of the pure-list items, and better designs should be explored by future studies. In Chapter 3, studying List 2 using the same encoding technique as List 1 (i.e., reading aloud versus silently) may have influenced the effectiveness of using production as a contextual cue at retrieval. A critical follow-up would be to investigate possible changes in the pattern of results if participants used a different encoding strategy when studying List 2, which might occur in typical list-method directed forgetting paradigms after participants receive an instruction to forget (Sahakyan & Delaney, 2003). In Chapter 4, when two lists of silent items were studied within the same trial, memory of List 1 was better than List 2—this was an unexpected finding that requires future verification.

Apart from these specific limitations, I would note that an informative extension to support the present findings would be the use of EEG in examining the brain processes that relate to the production effect. An EEG setup measures electrical activity within the brain using electrodes placed at various regions on the scalp; the associated brain activity is represented as event-related potentials (ERPs) as a function of time from the presentation of the stimulus of interest. To date, only one preliminary study has examined ERP evidence specifically for the production effect: Hassall et al. (2016) observed an increased amplitude in the “P300” ERP component over the parietal scalp regions in response to instructions to read aloud versus silently during encoding. The P300 has previously been related to distinctive processing and enhanced memory for studied items (e.g., Karis et al., 1984; Fabiani & Donchin, 1995; Kamp et al., 2012):

For example, Fabiani and Donchin (1995) showed that in the von Restorff effect (von Restorff, 1933), physically (i.e., font size) and semantically distinct words evoked greater P300 amplitude during encoding than did non-distinct words, and that this was associated with better memory for the distinct words. The P300 is thus a plausible index for the distinctive processing of aloud relative to silent words—and by extension, the formation of a distinct contextual framework—in a production effect paradigm. A critical investigation into brain activity during the *test phase* of the production effect paradigm is called for to show that participants indeed use production as a contextual cue to facilitate retrieval.

Related work has reported ERP signal differences when testing recognition for items that have a relatively stronger source memory. For example, in the remember/know paradigm (e.g., Gardiner, 1988), participants make a “remember” judgment when they can recollect contextual information associated with studying a particular item (e.g., any distinct visual or auditory information), and a “know” judgment when the studied item feels familiar but they do not recollect any contextual information. A “remember” judgment is thus related to having source memory for the studied item. Duzel et al. (1997) showed an enhanced “late positive” ERP component from the temporoparietal regions of the scalp for words that participants indicated that they “remembered” relative to the words that they simply “knew.” Woroch and Gonsalves (2010) provided another example using studied computer-generated visual objects: Participants showed enhanced late positive ERP signals from the parietal regions for correct relative to incorrect source judgments at test, and this was further associated with participants’ self-ratings of source memory strength. This difference was not observed for an earlier time window, where enhanced activity emerging from the frontal regions was associated with item memory for the

studied objects (corresponding to the “FN400”); this component is thought to reflect an index of memory without regard for recollection of source details.

Based on these related findings, when testing for recognition memory in the production effect, the following dissociation should occur at test. There should be enhanced P300 and/or late positive ERP components (i.e., increased amplitudes) measured from the parietal scalp regions for correctly recognized produced items relative to unproduced items. In contrast, because it does not distinguish source memory, the FN400 component should not differentiate between correctly recognized produced items and unproduced items. Such neural evidence for the processing of source memory information would strengthen the central aim of this dissertation for characterizing the production effect as a context-based memory effect. An ERP study to test this hypothesis was initiated for this dissertation but the COVID-19 pandemic prevented sufficient data collection.

Conclusion

In summary, the three studies constituting this dissertation advance the examination of the distinctiveness property of studying items aloud versus silently as a context-based memory effect, both when the test is recognition and when the test is recall. In Study 1, where the test was recognition, this distinctiveness property was useful in enhancing memory even when only aloud items were studied, suggesting that participants could retrieve this relevant contextual cue for the aloud items by contrasting them with the non-studied distractors. In Study 2, the contextual change account of directed forgetting was used to show that the “aloud” information indeed represented the relevant contextual information that is encoded in a production effect paradigm: Instructing people to use this information helped them overcome the effects of directed forgetting. Finally, in Study 3, the distinctiveness aspect of studying items aloud was examined

within recall tests, showing that this property enhanced memory for aloud items by interfering with memory of previously studied items. This was attributed to studying items aloud causing a release from buildup of proactive interference from earlier items due to a change of context.

In all of these situations, reading information aloud served as a useful context-based memory technique that is applicable to improving memory for key information. The findings presented in this dissertation shows that production is a simple-to-apply yet effective contextual cue that is easily attached to information when it is read aloud. At retrieval, production reliably acts as a source cue—as a context—clearly enhancing memory beyond simply learning by rote memorization. When there is limited opportunity to study information and it is not practical to use more elaborate encoding techniques, reading aloud is perhaps the most straightforward technique for boosting one's learning.

References

- Abel, M., & Bäuml, K.-H. T. (2019). List-method directed forgetting after prolonged retention interval: Further challenges to contemporary accounts. *Journal of Memory and Language, 106*, 18-28.
- Algina, J., & Keselman, H. J. (2003). Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement, 63*, 537-553.
- Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork, & R. A. Bjork (Eds.), *Memory* (pp. 237-313). Academic Press.
- Barlow, M. C. (1928). The role of articulation in memorizing. *Journal of Experimental Psychology, 11*, 306-312.
- Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 603-616.
- Bauml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language, 68*, 39-53.
- Benjamin, A. S. (2006). The effects of list-method directed forgetting on recognition memory. *Psychonomic Bulletin & Review, 13*, 831-836.
- Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition, 5*, 176-196.
- Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9*, 255-268.
- Bodner, G. E., & MacLeod, C. M. (2016). The benefits of studying by production... and of studying production: Introduction to the special issue on the production effect in memory.

- Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 70, 89-92.
- Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1711-1719.
- Bodner, G. E., 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, 93-98.
- Bransford, D. J., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717-726.
- Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107, 127-181.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106, 551-581.
- Chee, Q. W., & Goh, W. D. (2018). What explains the von Restorff effect? Contrasting distinctive processing and retrieval cue efficacy. *Journal of Memory and Language*, 99, 49-61.
- Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, 22, 267-281.
- Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, 26, 341-361.
- Crowder, R. G. (1976). *Principles of learning and memory*. Lawrence Erlbaum.

- Cumming, G. (2012). *Understanding the new statistics*. Routledge.
- Dallett, K. M., & Wilcox, S. G. (1968). Contextual stimuli and proactive inhibition. *Journal of Experimental Psychology*, *78*, 475-480.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.
- 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*, 155-161.
- DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift? *Current Opinion in Behavioral Sciences*, *17*, 141-146.
- Duzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H.-J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences*, *94*, 5973-5978.
- Earhard, M. (1967). Cued recall and free recall as a function of the number of items per cue. *Journal of Verbal Learning and Verbal Behavior*, *6*, 257-263.
- Eich, J. E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition*, *8*, 157-173.
- Eldridge, L. L., Engel, S. A., Zeineh, M. M., Bookheimer, S. Y., & Knowlton, B. J. (2005). A dissociation of encoding and retrieval processes in the human hippocampus. *Journal of Neuroscience*, *25*, 3280-3286.
- Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 671-682.

- Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 224-240.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.
- Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*, 1-5.
- Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*, 99-115.
- Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science*, *27*, 302-308.
- 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*, 125-138.
- Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (2014). Production improves memory equivalently following elaborative vs non-elaborative processing. *Memory*, *22*, 470-480.
- Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*, 1046-1055.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309-313.

- Gardiner, J. M., Craik, F. I., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, *11*, 778-783.
- Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*, 110-119.
- Godden, D., & Baddeley, A. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*, 325-331.
- Guitard, D., Saint-Aubin, J., & Cowan, N. (2021). Asymmetrical interference between item and order information in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*, 243-263.
- 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*, 139-146.
- Henson, R. N. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*, 73-137.
- Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 503-508.
- Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 484-494.
- Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*, 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, 242-246.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269-299.
- Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, 2, 105-112.
- Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science*, 22, 10-15.
- Icht, M., & Mama, Y. (2019). The effect of vocal production on vocabulary learning in a second language. *Language Teaching Research*.
- Icht, M., Mama, Y., & Algom, D. (2014). The production effect in memory: Multiple species of distinctiveness. *Frontiers in Psychology*, 5, 886.
- Isarida, T., & Isarida, T. K. (2007). Environmental context effects of background color in free recall. *Memory & Cognition*, 35, 1620-1629.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513-541.
- Jamieson, R. K., & Spear, J. (2014). The offline production effect. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 68, 20-28.
- Jamieson, R. K., Mewhort, D. J., & 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, 154-164.
- JASP Team. (2020). *JASP* (Version 0.14)[Computer software].

- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.
- 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*, 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*, 441-448.
- Jonker, T. R., Seli, P., & MacLeod, C. M. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review, 120*, 852-872.
- Jonker, T. R., Seli, P., & MacLeod, C. M. (2015). Retrieval-induced forgetting and context. *Current Directions in Psychological Science, 24*, 273-278.
- Kamp, S.-M., Forester, G. R., Murphy, A. R., Brumback, T., & Donchin, E. (2012). Testing a distinctiveness explanation of the primacy effect in free recall using event-related potentials. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, (pp. 539-544). Sapporo, Japan.
- Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology, 16*, 177-216.
- Kelley, K. (2007). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods, 39*, 979-984.
- Kelly, M. O., Ensor, T. M., Lu, X., MacLeod, C. M., & Risko, E. F. (in press). Reducing retrieval time modulates the production effect: Empirical evidence and computational accounts. *Journal of Memory and Language*.

- Kliegl, O., & Bäuml, K.-H. T. (2021). Buildup and release from proactive interference - Cognitive and neural mechanisms. *Neuroscience and Biobehavioral Reviews*, *120*, 264-278.
- Kurtz, K. H., & Hovland, C. I. (1953). The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *Journal of Experimental Psychology*, *45*, 157-164.
- 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*, 165-176.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 970-988.
- Lehman, M., & Malmberg, K. J. (2011). Overcoming the effects of intentional forgetting. *Memory & Cognition*, *39*, 335-347.
- Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 1-11.
- 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*, 212-216.
- Loft, S., Humphreys, M. S., & Whitney, S. J. (2008). Control of access to memory: The use of task interference as a behavioral probe. *Journal of Memory and Language*, *58*, 465-479.

- MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98*, 291-310.
- MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding, & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1-57). Lawrence Erlbaum Associates.
- MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26*, 390-395.
- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 671-685.
- Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112*, 122-133.
- Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35*, 545-556.
- Malmberg, K. J., Lehman, M., & Sahakyan, L. (2006). On the cost and benefit of taking it out of context: Modeling the inhibition associated with directed forgetting. *Proceedings of the 28th Meeting of the Cognitive Science Society*, (pp. 549-554). Vancouver, BC.
- McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15*, 237-255.
- McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 563-582.
- McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39*, 352-370.

- Mensink, G.-J. M., & Raaijmakers, J. G. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, *33*, 172-186.
- Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434-455.
- Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*, 638-677.
- Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*, 403-415.
- Muther, W. S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science*, *3*, 429-430.
- Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, *16*, 343-352.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251-269.
- Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 702-709.
- Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1543-1547.
- Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*, 326-338.

- Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory, 20*, 717-727.
- Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory, 22*, 509-524.
- Paivio, A. (1971). Imagery and language. In Segal, & S. Joelson (Eds.), *Imagery: Current cognitive approaches* (pp. 7-32). Academic Press.
- Pastötter, B., & Bäuml, K.-H. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 54-65.
- Pastötter, B., Bäuml, K.-H., & Hanslmayr, S. (2008). Oscillatory brain activity before and after an internal context change - Evidence for a reset of encoding processes. *NeuroImage, 43*, 173-181.
- Pastötter, B., Kliegl, O., & Bäuml, K.-H. T. (2012). List-method directed forgetting: The forget cue improves both encoding and retrieval of postcue information. *Memory & Cognition, 40*, 861-873.
- Pastötter, B., Kliegl, O., & Bäuml, K.-H. T. (2016). List-method directed forgetting: Evidence for the reset-of-encoding hypothesis employing item-recognition testing. *Memory, 24*, 63-74.
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 287-297.
- Pastötter, B., Tempel, T., & Bäuml, K.-H. T. (2017). Long-term memory updating: The reset-of-encoding hypothesis in list-method directed forgetting. *Frontiers in Psychology, 8*, 2076.

- Perfect, T. J., Mayes, A. R., Downes, J. J., & van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology*, *49A*, 797-813.
- Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*, 904-915.
- Rutherford, A. (2004). Environmental context-dependent recognition memory effects: An examination of ICE model and cue-overload hypotheses. *The Quarterly Journal of Experimental Psychology*, *57A*, 107-127.
- Sahakyan, L., & Delaney, P. F. (2003). Can encoding differences explain the benefits of directed forgetting in the list method paradigm? *Journal of Memory and Language*, *48*, 195-206.
- Sahakyan, L., & Delaney, P. F. (2005). Directed forgetting in incidental learning and recognition testing: Support for a two-factor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 789-801.
- Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1064-1072.
- Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 59, pp. 131-189). Academic Press.
- Sahakyan, L., Waldum, E. R., Benjamin, A. S., & Bickett, S. P. (2009). Where is the forgetting with list-method directed forgetting in recognition? *Memory & Cognition*, *37*, 464-476.

- Shin, Y. S., Masís-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2021). Context-dependent memory effects in two immersive virtual reality environments: On Mars and underwater. *Psychonomic Bulletin & Review*, *28*, 574-582.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592-604.
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 460-471.
- Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition*, *10*, 405-412.
- Smith, S. M. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory & Cognition*, *12*, 477-482.
- Smith, S. M., & Manzano, I. (2010). Video context-dependent recall. *Behavior Research Methods*, *42*, 292-301.
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203-220.
- Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342-353.
- Smithson, M. (2003). *Quantitative applications in the social sciences: Confidence intervals*. SAGE Publications, Inc.
- Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*, 164-182.

- Szpunar, K. K., McDermott, K. B., & Roediger, H. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1392-1399.
- Tan, P., Ensor, T. M., Hockley, W. E., Harrison, G. W., & Wilson, D. E. (2020). In support of selective rehearsal: Double-item presentation in item-method directed forgetting. *Psychonomic Bulletin & Review*, *27*, 529-535.
- Taylor, T. L., Cutmore, L., & Pries, L. (2018). Item-method directed forgetting: Effects at retrieval? *Acta Psychologica*, *183*, 116-123.
- Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology*, *88*, 248-257.
- Thomas, R. C., & McDaniel, M. A. (2013). Testing and feedback effects on front-end control over later retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 437-450.
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., & Spooner, P. A. (2007). Schemas and memory consolidation. *Science*, *316*, 76-82.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352-373.
- Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior*, *13*, 181-193.
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*, 49-60.

- Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Focusing the search: Proactive and retroactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1742-1756.
- Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Evidence for noisy contextual search: Examining the dynamics of list-before-last recall. *Memory*, *20*, 1-13.
- van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*, 813-826.
- von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologie Forschung*, *18*, 299-342.
- Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, *63*, 410-424.
- Wammes, J. D., Roberts, B. R., & Fernandes, M. A. (2018). Task preparation as a mnemonic: The benefits of drawing (and not drawing). *Psychonomic Bulletin & Review*, *25*, 2365-2372.
- Watkins, M. J., & Sechler, E. S. (1988). Generation effect with an incidental memorization procedure. *Journal of Memory and Language*, *27*, 537-544.
- Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *104*, 442-452.

- Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1024-1039.
- Woroch, B., & Gonsalves, B. D. (2010). Event-related potential correlates of item and source memory strength. *Brain Research*, *1317*, 180-191.
- Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition*, *35*, 526-544.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517.
- Zhou, Y., & MacLeod, C. M. (2021). Production between and within: Distinctiveness and the relative magnitude of the production effect. *Memory*, *29*, 168-179.

Appendix A

List of words used for Chapter 2. The words assigned to each condition (aloud, silent, distractor) were always selected randomly from the entire set of words for each participant in every experiment.

Experiment 1

account	address	afternoon	amount	answer	arrow
attention	attitude	author	avenue	basket	battery
beauty	border	branch	building	campaign	capital
captain	castle	century	clothes	daughter	debate
department	dinner	direction	distance	education	election
engine	entrance	envelope	evening	factory	fashion
forest	foundation	friend	furniture	garden	glass
gravity	guardian	handle	harbour	history	holiday
industry	invention	invitation	island	journey	judge
justice	kettle	kingdom	kitchen	knock	ladder
language	laugh	leather	lesson	machine	market
meadow	merchant	message	minute	neighbour	nephew
ocean	office	orchard	package	painting	partner
peace	pebble	plate	pocket	porch	powder
quarrel	quarter	queen	record	resort	reward
river	sailor	school	shadow	shoulder	speech
station	steam	stream	summer	teacher	theatre
thread	ticket	traffic	travel	treasure	trousers
turnip	uncle	uniform	vacation	valley	victory
village	wagon	wheat	wheel	whisper	winter

Additional words used in Experiment 2

article	automobile	battle	block	blossom	bottle
breakfast	bridge	business	character	circle	citizen
coast	cotton	creature	crowd	curtain	desire
doctor	dream	empire	expense	family	feather
fence	figure	fortune	frame	frost	governor
guest	harvest	health	interest	knight	knowledge
league	library	match	material	member	moment
mountain	nation	notice	object	occasion	opinion
penny	permit	plain	prince	property	province
purse	region	result	review	ribbon	robin
saddle	season	section	sense	service	shape
shell	shelter	shore	society	space	stairs
stamp	street	string	surface	temple	vessel
window	witness				

Additional words used in Experiment 2 and removed in Experiment 3:

banner	barrel	breath	breeze	bucket	channel
charity	colony	column	compass	committee	continent
council	discovery	display	division	flight	information
instruction	method	monarch	monument	motion	passenger
pattern	policy	population	portion	position	principle
prospect	receipt	relation	reserve	slipper	source
sunshine	territory	vision	voyage		

Appendix B

The list of words used for Chapter 3. The word presentation order and the assignment of words to conditions were randomized anew for each participant.

account	address	afternoon	amount	answer	arrow
article	attention	attitude	author	automobile	avenue
banner	barrel	basket	battery	battle	beauty
block	blossom	border	bottle	branch	breakfast
breath	breeze	bridge	bucket	building	business
campaign	capital	captain	castle	century	channel
character	charity	circle	citizen	clothes	coast
colony	column	committee	compass	continent	cotton
council	creature	crowd	curtain	daughter	debate
department	desire	dinner	direction	discovery	display
distance	division	doctor	dream	education	election
empire	engine	entrance	envelope	evening	expense
factory	family	fashion	feather	fence	figure
flight	forest	fortune	foundation	frame	friend
frost	furniture	garden	glass	governor	gravity
guardian	guest	handle	harbour	harvest	health
history	holiday	industry	information	instruction	interest
invention	invitation	island	journey	judge	justice
kettle	kingdom	kitchen	knight	knock	knowledge
ladder	language	laugh	league	leather	lesson
library	machine	market	match	material	meadow
member	merchant	message	method	minute	moment
monarch	monument	motion	mountain	nation	neighbour
nephew	notice	object	occasion	ocean	office
opinion	orchard	package	painting	partner	passenger
pattern	peace	pebble	penny	permit	plain
plate	pocket	policy	population	porch	portion
position	powder	prince	principle	property	prospect

province	purse	quarrel	quarter	queen	receipt
record	region	relation	reserve	resort	result
review	reward	ribbon	river	robin	saddle
sailor	school	season	section	sense	service
shadow	shape	shell	shelter	shore	shoulder
slipper	society	source	space	speech	stairs
stamp	station	steam	stream	street	string
summer	sunshine	surface	teacher	temple	territory
theatre	thread	ticket	traffic	travel	treasure
trousers	turnip	uncle	uniform	vacation	valley
vessel	victory	village	vision	voyage	wagon
wheat	wheel	whisper	window	winter	witness

Appendix C

The list of 186 words used in Chapter 4. The word presentation order was randomized anew for every participant.

amount	angle	answer	atom	author	back
beach	beauty	better	blame	blue	bond
bound	bread	breath	bride	bridge	camp
campus	career	cast	catch	cause	chair
chance	city	class	clay	coat	cold
column	core	cotton	count	crisis	crowd
deal	debate	decade	deep	degree	design
dollar	door	draw	drove	dust	duty
east	editor	effort	empty	escape	even
event	extra	factor	feel	fellow	felt
figure	film	firm	folk	former	fruit
future	gain	gift	glance	grave	group
hair	hand	heart	hero	hold	horse
hotel	island	issue	jacket	jazz	keep
last	lawyer	lead	leader	leave	light
like	list	look	loose	lord	luck
method	middle	mile	mine	miss	moment
mood	moon	motion	native	navy	number
order	paper	part	pass	past	permit
pilot	plan	play	poetry	pride	rain
rear	regard	result	river	rock	roof
room	safety	scene	seat	senate	serve
shadow	shape	share	sheet	shop	sight
skill	skin	soft	space	speech	speed
state	stone	store	story	strike	summer
supper	survey	table	talk	tape	target
task	temple	text	thank	thick	thing
till	tone	total	trade	train	type

under	valley	value	voice	wait	wall
want	warm	week	weight	well	wheel
wine	winter	wise	wish	wrong	youth