CHAPTER FOUR

List-Method Directed Forgetting in Cognitive and Clinical Research: A Theoretical and Methodological Review

Lili Sahakyan[1], Peter F. Delaney, Nathaniel L. Foster, Branden Abushanab
Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC, USA
[1]Corresponding author: E-mail: l_sahaky@uncg.edu

Contents

1. Introduction 133
2. List-Method DF: Design and Measurement 134
   2.1. Basic Design Issues 134
   2.2. Multilist Designs in DF 136
   2.3. Additional Dependent Measures in DF 137
3. Our Framework of List-Method DF 138
   3.1. Older Ideas about the Causes of DF 138
   3.2. A Two-Factor, Process-Based Account 139
4. Forgetting is a Strategic Decision 140
   4.1. Importance of Controlled Strategies 141
   4.2. Why do Some People Fail to Engage in DF? 142
   4.3. Beliefs about Memory Developed during the Experiment 143
   4.4. Preexisting Beliefs about Intentional Forgetting 144
   4.5. Forgetting Strategies and Their Relative Effectiveness 146
5. Context Change as an Explanation for DF Impairment 148
   5.1. Mental Context-Change Paradigm and Its Relation to DF 149
   5.2. Recognition Tests and Indirect Tests of Memory 150
   5.3. Mental Reinstatement of Context 153
   5.4. Cuing with Context Cues vs Category Cues 154
   5.5. Encoding Strength: Item vs Context Strength 155
   5.6. Contextual Variability 156
   5.7. Extralist Cued Recall 157
   5.8. Importance of L2 Learning 158
   5.9. Related Items Across Lists 159
   5.10. Retrieving L2 before L1 160
   5.11. Delay Effects 161
   5.12. Individual Differences in DF as a Function of Working Memory Capacity 162
   5.13. Reexposure or Retrieval of L1 Items Affects Subsequent Memory 162
   5.14. Summary 164

Psychology of Learning and Motivation, Volume 59
ISSN 0079-7421, http://dx.doi.org/10.1016/B978-0-12-407187-2.00004-6

6. Areas of Disagreement Across Studies 165
6.1. Serial Position Effects 165
6.2. List Differentiation in DF: Source vs Intrusion Errors 166
7. Strategy Change Explains DF Benefits 167
7.1. The Advent of the Dual-Process Account 167
7.2. A Large-Scale Replication Study 169
7.3. Test Order and the Benefits of DF 170
7.4. Summary 170
8. Implications for Clinical Populations 172
8.1. Reinterpreting Clinical List-Method DF Results as Reflecting Context 173
8.1.1. Attention-Deficit Hyperactivity Disorder 173
8.1.2. Schizophrenia 174
8.1.3. Generalized Anxiety Disorder 175
8.1.4. Depression 175
8.2. Future Directions for Clinical Research Using the List Method 176
8.3. Improving Research Design for Clinical Studies of List-Method DF 178
8.3.1. Inclusion of a Separate Remember Condition 178
8.3.2. Controlling Test Order and Measuring Output Order 179
8.3.3. Scaling Effects 180
8.3.4. Floor Effects 180
8.4. Summary and Linking Clinical to Nonclinical Research on DF 181
9. Concluding Thoughts 181
Acknowledgments 182
References 182

Abstract

The primary purpose of this chapter is to provide an up-to-date review of the twenty-first century research and theory on list-method directed forgetting (DF) and related phenomena like the context-change effect. Many researchers have assumed that DF is diagnostic of inhibition, but we argue for an alternative, noninhibitory account and suggest reinterpretation of earlier findings. We first describe what DF is and the state of the art with regard to measuring the effect. Then, we review recent evidence that brings DF into the family of effects that can be explained by global memory models. The process-based theory we advocate is that the DF impairment arises from mental context change and that the DF benefits emerge mainly but perhaps not exclusively from changes in encoding strategy. We review evidence (some new to this paper) that strongly suggests that DF arises from the engagement of controlled forgetting strategies that are independent of whether people believed the forget cue or not. Then we describe the vast body of literature supporting that forgetting strategies result in contextual change effects, as well as point out some inconsistencies in the DF literature that need to be addressed in future research. Next, we provide evidence—again, some of it new to this chapter—that the reason people show better memory after a forget cue is that they change encoding strategies. In addition to reviewing the basic research with healthy population, we reinterpret the evidence from the literature on certain clinical populations, providing a critique of the work done to date and outlining ways of improving the methodology for the study of DF in special populations. We conclude with a critical discussion of alternative approaches to understanding DF.

1. INTRODUCTION

Forgetting is often a passive process that happens regardless of our conscious intentions. However, this is not always the case; forgetting sometimes is an active process of intentionally engaging strategies that reduce memory for material that interferes with one's current goals. To this end, forgetting can be an adaptive process. For example, to use software that has been recently updated, the user must forget the prior ways of doing things to avoid errors. Similarly, we are often told "never mind, forget what I just said" following exposure to erroneous information.

Intentional forgetting has been studied extensively in the laboratory via directed forgetting (DF) manipulations, which produce impaired memory for material following an instruction to try to forget. In DF studies, participants study some information and are subsequently told to forget certain portions of it. The *list-method* instructs people to forget an entire list of earlier studied items, whereas *the item-method* instructs people to forget or remember on item-by-item basis. Both methods demonstrate forgetting of unwanted information on demand, although the way by which forgetting is accomplished differs between procedures.

This chapter primarily centers on list-method DF, partly because the mechanisms proposed to explain it have been a topic of active debate in the recent decade, and partly because the presumed connection with inhibitory processes has made the paradigm an attractive diagnostic tool for investigating clinical and special populations. We provide an up-to-date review of DF and related phenomena (e.g. the context-change paradigm) with a focus on studies conducted since 2000. There is already an excellent edited volume describing twentieth century DF research (Golding & MacLeod, 1998).

We will begin by describing methodological issues in list-method DF. The designs used to study DF in cognitive research have evolved to keep pace with theoretical and methodological developments, and so we will review the difficulties posed by earlier designs and suggest better-controlled procedures. We will then outline our broad theory of DF and describe the large number of studies supporting that broad theory, as well as areas of current controversy in DF. Even researchers who are familiar with our previous manifestos on this topic may find this section interesting, as our view has evolved and expanded even since 2010, and because we will present previously unpublished data that bear on a number of issues. Finally, we will explore how the theory can inform studies of individual differences, especially in clinical populations, and make recommendations for conducting such studies.

2. LIST-METHOD DF: DESIGN AND MEASUREMENT

2.1. Basic Design Issues

DF is generally assessed in terms of two separate components of the effect: the costs and the benefits. To understand these terms, we first need to explain how most DF studies are set up, and then we will explain the usual findings.

The most frequently used design is the *two-list design*, which presents two lists of items (usually words) to study for a later memory test. The first list we will call L1 and the second list we will call L2. Following L1, a *cue* is presented either to keep remembering L1 or to forget L1. Because in the two-list design the cue always occurs between L1 and L2, we sometimes refer to L1 as the precue items and L2 as the postcue items, which is helpful when considering designs that include more lists (see below for a discussion of these designs). There are a variety of ways to present the forget cue, but a typical instruction is to try to forget L1 because it was "just for practice". The remember cue provides a control group against which to assess the effects of the forget instruction, and it typically involves telling people to keep remembering L1 because it was only the first half of the study material. A filler task is usually inserted after learning L2 to reduce recency effects. Afterward, participants receive a recall test for both lists, including items that they were told to forget. In free recall, the forget instruction reduces recall of L1 items relative to the remember group—a phenomenon called the *costs* of DF. Similarly, the forget instruction produces better memory of L2 items relative to the remember group—a phenomenon called the *benefits* of DF (for reviews, see Bäuml, 2008; Bjork, Bjork, & Anderson, 1998; MacLeod, 1998).

Whether participants are instructed to recall L1 or L2 first varies across studies. Some studies have required participants to recall L1 first and then recall L2 (e.g. Delaney & Sahakyan, 2007; Foster & Sahakyan, 2011; Mulji & Bodner, 2010; Pastötter & Bäuml, 2007; Spillers & Unsworth, 2011). Other studies have fully counterbalanced the test orders (e.g. Kimball & Bjork, 2002; Lehman & Malmberg, 2009, 2011a; Pastötter, Kliegl, & Bäuml, 2012; Sahakyan & Delaney, 2010; Zellner & Bäuml, 2006), and some studies have allowed participants to recall all items in any order (Golding & Gottlob, 2005; Joslyn & Oakes, 2005; Minnema & Knowlton, 2008; Wessel & Merckelbach, 2006; Zellner & Bäuml, 2006).

One obvious concern is that differences in output order might affect DF through output interference (Anderson, 2005). Output interference is the finding that retrieving some items from a set first reduces the likelihood of retrieving the rest of the set (Dong, 1972; Roediger, 1974; Roediger & Schmidt, 1980; Tulving & Arbuckle, 1966). In DF, when test order is not controlled people might be inclined to begin recall with L2 items due to recency, causing output interference on L1 items. Given that L2 is better remembered in the forget group than the remember group, there may be greater output interference on L1 in the forget group. Indeed, when left to recall lists in any order, participants prefer to initiate recall with postcue remember items, thereby resulting in an effect of output order (Golding & Gottlob, 2005). However, a recent meta-analysis of list-method DF studies that included both forget and remember groups did not support this notion; initiating retrieval with L2 did not exaggerate the magnitude of DF costs, although it made it easier to detect the DF benefits (Pastötter et al., 2012). Although these results suggest that output order affects the benefits more than the costs (at least with unrelated items across the lists), we nevertheless suggest fully controlling the test order of each list to obtain more "pure" measures of recall uncontaminated by prior retrieval. Output interference could become a particularly thorny issue if the lists contain a mixture of items (e.g. neutral and emotional words), which could create different degrees of output interference if the test order is left uncontrolled.

Another procedural inconsistency across DF studies is whether a control group is included in the design. In the past, many DF paradigms have failed to include a control group wherein both L1 and L2 presentations are followed by a remember cue. In such paradigms, the magnitude of forgetting was quantified by comparing recall performance for the forget list (L1) relative to the remember list (L2). This measure is referred to as the "R–F" measure. The justification for not including a control group is based on the notion that the only ostensible difference between L1 and L2 is the type of cue received. However, there are several confounding variables that highlight the necessity for a control group. First, heightened recall performance for L2 items could be attributed to a recency effect, making it difficult to assume that L2 gains are a result of L1 forgetting. Second, increased accessibility for L2 items may be attributed to learning effects. That is, following L1 exposure, participants may be more adapted to the procedure, leading them to shift their encoding strategies for L2 items. Finally, reduced recall performance for L1 items may be attributed to retroactive interference accrued from L2 items. In other words, at the time of recall, L1 accessibility is reduced due to the interference built up from L2 encoding. Given the number of alternative explanations for reduced L1 recall and heightened L2 recall, it is challenging to attribute these findings solely to DF without comparing recall performance to a control group that was required to remember both lists.

2.2. Multilist Designs in DF

A variation on the traditional two-list procedure is the use of a *three-list design*, which was put forth to reduce potential confounds in the two-list procedure (Lehman & Malmberg, 2009). The sources of these confounds are as follows. First, L2 presentation is typically followed by a filler task, whereas L1 presentation is not, which means that L2 memory could be penalized in comparison with L1. For example, participants can covertly rehearse the end of L1 during the beginning of L2 encoding, increasing memory for the last few L1 items, while hindering encoding of the first few L2 items. Additionally, in a traditional two-list design, L2 is subjected to proactive interference from L1, but L1 receives no interference from a prior list. Therefore, the two-list approach may benefit L1 and harm L2. To control these factors, in a three-list design, participants study three lists, each followed by a brief filler activity of equal duration before a forget or a remember cue is presented. For half of the participants, the forget cue is presented after L2, whereas the remaining half are told to remember all lists. At the time of final test, half of the participants are required to recall either L2 or L3 (L1 is never tested). Note that the three-list design could lead to potential floor effects on L2 if presentation rates are not substantially increased (e.g. Lehman & Malmberg, 2011a). Both the two-list and the three-list designs manipulate the forget/remember cue between-subjects.

An alternative approach is the use of a *four-list design*, where the forget/remember cue is varied within-subjects (e.g. Zellner & Bäuml, 2006). In this procedure, two lists are assigned to the first block, and two lists are assigned to the second block. The forget cue occurs either during the first block (e.g. after presentation of L1) or during the second block (e.g. after presentation of L3). Participants complete the forget condition (F, R) for the first block and the remember condition (R, R) for the second block, or vice versa. Memory for both lists is tested after each block, with counterbalanced test order of the lists. This procedure produces findings similar to the traditional two-list version. One caveat concerns a potential shift to better encoding strategies after the first block has been completed and memory for both lists has been tested (Delaney & Knowles, 2005; Sahakyan & Delaney, 2003). With that in mind, there may be some order effects depending on whether forgetting is inferred from the RR–FR order or the FR–RR order. For example, the size of the DF impairment may be exaggerated in the FR–RR order because by the time of L3, participants may adopt better encoding strategies on L3. Although Bäuml et al. typically do not obtain session order effects, faster presentation rates that are employed in some of their studies may mask such effects because with faster rates there is leaving little room for employing a better encoding strategy than rehearsal. An order effect has been reported in one study that used a slow presentation rate and obtained larger DF when the forget cue occurred in the first block compared to the third block (Hanczakowski, Pasek, & Zawadska, 2012). Therefore, researchers should be mindful of potential order effects that may arise from use of the four-list, within-subjects approach to DF.

2.3. Additional Dependent Measures in DF

DF has been assessed primarily in terms of the number of correctly recalled or recognized items from each list (e.g. costs/benefits). In its early days, it was inferred through an even less-sensitive measure (e.g. "R–F" difference). However, recently researchers have strived to include a wider array of dependent measures with the goal of constraining the interpretations. Such measures include investigations of serial position functions (Geiselman, Bjork, & Fishman, 1983; Lehman & Malmberg, 2009; Pastötter & Bäuml, 2010; Pastötter et al., 2012; Sahakyan & Foster, 2009; Sheard & MacLeod, 2005), intrusion errors (Lehman & Malmberg, 2009; Sahakyan & Delaney, 2010; Spillers & Unsworth, 2011), recall latencies, which indicate an average time point in the recall period when responses were emitted (Spillers & Unsworth, 2011; Unsworth, Spillers, & Brewer, 2012), first response functions, which provide a measure of where in the list participants initiate their recall (e.g. Lehman & Malmberg, 2009), and conditional response probabilities, which indicate how people transition between responses during recall (e.g. Unsworth et al., 2012). The use of multiple measures provides a more detailed assessment of the recall process, thereby facilitating a deeper understanding of the mechanisms underlying list-method DF. Together, they coalesce to provide a broader perspective on list-method DF, which in turn can be used to guide and advance future research.

3. OUR FRAMEWORK OF LIST-METHOD DF

3.1. Older Ideas about the Causes of DF

The earliest theory of DF was the *selective rehearsal account* (e.g. Bjork, 1970, 1972), which was proposed at the time when not much was known about the distinction between the list-method and item-method DF. According to this view, the costs arise because participants stop rehearsing to-be-forgotten items in response to the forget cue, and they devote their rehearsal and mnemonic activity more effectively to to-be-remembered items, producing the benefits. This works well in explaining item-method DF, but its suitability for explaining list-method DF started to wane when it became clear that the list-method and item-method were differentially sensitive to recognition and indirect tests of memory (e.g. Basden, Basden, & Gargano, 1993; MacLeod, 1999). Additionally, the presence of significant DF in incidental learning (Geiselman et al., 1983; Sahakyan & Delaney, 2005, 2010; Sahakyan, Delaney, & Goodmon, 2008) made it increasingly difficult for the selective rehearsal account to encompass list-method DF, where it was soon supplemented by the *retrieval inhibition* account (Bjork, 1989; Geiselman et al., 1983).

According to the retrieval inhibition account, when participants are instructed to forget L1, they initiate an inhibitory process that suppresses or deselects that list so as to facilitate the learning of subsequent lists. Therefore, L1 memory suffers from inhibition, and L2 memory benefits because inhibited L1 items do not cause proactive interference on L2.

The term *inhibition* is interpreted differently among DF researchers. Some interpret inhibition as suppressing the activation level of L1 items (e.g. Barnier, Conway, Mayoh, Speyer, Avizmil, & Harris, 2007; Conway, Harries, Noyes, Racsmány, & Frankish, 2000; Racsmány & Conway, 2006; Racsmány et al., 2008); others propose that items reside in memory at their full strength as evidenced by indices of memory other than free recall, but retrieval of the L1 episode is inhibited (e.g. Bjork, 1989; Bjork & Bjork, 1996), and recently some have argued that the L1 context may be inhibited (e.g. Anderson, 2005; Bäuml, 2008; Pastötter & Bäuml, 2010). Multiple meanings of the term *inhibition* can be misleading, especially to researchers outside of the memory field who employ the DF paradigm as a test of inhibitory function in various populations. Equating an empirical pattern of findings with an underlying inhibitory mechanism is unwarranted because mechanisms other than inhibition could lead to impaired access (e.g. context change and blocking).

Historically, in the mid-to-late 1980s, inhibition was used to explain the discrepant findings in item-method and list-method DF. However, nothing was assumed about the nature of inhibition or the conditions under which inhibition should be more or less evident. The unspecified nature of inhibition makes it difficult to think of it as a mechanism that explains DF as opposed to providing a description of empirical findings. Later, in the 1990s, inhibition was developed into a full theoretical mechanism explaining retrieval-induced forgetting (RIF) (e.g. Anderson, 2003; Anderson, Bjork, & Bjork, 1994), which is a memory impairment phenomenon that occurs at the item-level, unlike the list-method DF, which occurs at the list-level. Nowadays inhibition account makes testable predictions for RIF, but it fails to make predictions for DF because it is unclear what exactly is being inhibited in DF. Just about any outcome can be interpreted to be consistent with inhibition, which significantly reduces its appeal as a theory of DF. That DF is frequently reviewed together with other inhibitory paradigms (e.g. RIF and think-no-think procedure) may have contributed to inadvertent blending of the use of the term *inhibition* across paradigms (e.g. Anderson, 2003, 2005; Depue, 2012).

We take issue with the notion that the "L1 episode" is suppressed as a result of the forget cue. Someone not well-versed in the nuances of memory theory is likely to interpret "inhibition of L1 episode" as analogous to "inhibition of items within that episode". However, the L1 episode contains more than just the studied items within that list. It contains contextual information, which distinguishes that particular episode from other similar episodes (e.g. the L2 episode). Thus, an episode is a complex term that refers to both the items and the context within which those items were experienced. We argue that there is no evidence for inhibition of items, and we further argue that there is no evidence for inhibition of context. Instead, we explain the impaired accessibility to L1 items by invoking the notion of a mismatch between the retrieval cues being used to search the memory and the contents of memory.

3.2. A Two-Factor, Process-Based Account

Figure 4.1 shows how we currently understand list-method DF. Assuming a two-list design, the forget cue takes place after L1 and triggers two types of reflective processing. First, it triggers a decision about how to comply with the forget instruction. We argue that the primary mechanism producing forgetting involves a change of mental context between the two lists, and that some strategies (but not all) bring about this contextual change.

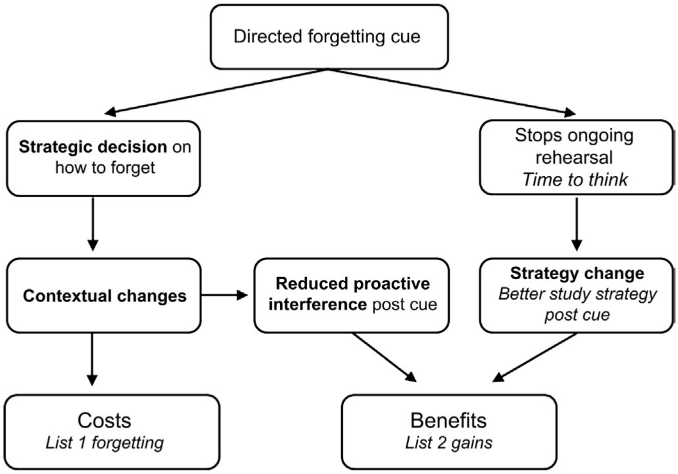

Figure 4.1 A two-factor, process-based framework for list-method directed forgetting.

Context change also contributes somewhat to the benefits by reducing proactive interference on L2, but this effect is often small and difficult to detect, particularly when there are only two lists. Second, the forget cue stops ongoing rehearsal processes on L1, which gives participants time to think about what they should do to learn L2. This leads some participants to change study strategies and consequently improve L2 learning. We think that this strategy-change component accounts for the improved memory for postcue material.

Sections 4–6 provide the rationale for each of these mechanisms and the empirical support for them. Section 4 discusses the importance of controlled strategies for obtaining DF, and how metacognitive beliefs affect the decision to deploy those strategies. Section 5 details the mental context-change account and Section 6 explains the strategy-change account. We will therefore wait to explain in detail why we think these mechanisms provide a better explanation for list-method DF phenomena than earlier accounts do.

4. FORGETTING IS A STRATEGIC DECISION

In a series of studies from our lab, we have explored what people do in response to the forget cue. Specifically, we have accumulated a wealth of information regarding how participants interpret and comply with forget instructions. People may think that they already have forgotten, and therefore fail to deploy any strategy to forget. Likewise, they may not know what kind of strategy to deploy in order to forget, and therefore do nothing.

Such results are important not only for demonstrating that forgetting is an active process that requires conscious activity in order to forget, but also for explaining individual differences—including between-group differences—in the ability to forget unwanted information.

4.1. Importance of Controlled Strategies

Many studies in our lab have revealed that obtaining DF requires that participants engage in controlled behaviors to reduce accessibility of unwanted information. For example, Foster and Sahakyan (2011) collected retrospective verbal reports regarding what participants did in response to the forget instruction, including whether they believed/trusted the forget cue. In a group of 80 forget-group participants, 25% reported "doing nothing", whereas the remaining participants reported a variety of strategies, which we collectively termed as "doing something". Nearly 34% of the forget-group participants reported "not believing" the forget cue. Surprisingly, however, there was virtually no correlation ($\varphi = 0.08$) between forget-cue belief and whether participants engaged in controlled strategies in order to forget. Some participants trusted the forget cue but did nothing in response to it, whereas others engaged in forgetting strategies despite not trusting the cue. It seemed like the latter group decided to "play along" despite having doubts about the truthfulness of the forget cue. Most importantly, L1 recall in the forget group was entirely determined by whether participants engaged in controlled strategies or not, whereas believing in the forget cue was completely irrelevant. When the forgetting strategy (doing something vs nothing) and the belief in the forget cue (yes vs no) were simultaneously entered as predictors of L1 recall in the forget group, only the strategy variable was a significant predictor; the belief variable was not. The results are summarized in Figure 4.2.

Instructions to forget in the real world are sometimes explicit. For example, if construction is blocking a geographical route, one must temporarily "forget the original route" in favor of a more contextually appropriate route (e.g. Golding & Keenan, 1985). However, an implicit instruction to forget may be just as relevant, such as forgetting a prior credit card number to facilitate memory for a recently issued card. We compared the effectiveness of explicit and implicit forget cues and showed that the magnitude of DF is unaffected by the nature of the forget instruction (Foster & Sahakyan, 2011). Regardless of whether forgetting comes from implicit needs of the environment, or explicit prompts of others, it leads to similar behavioral effects.

These findings highlight several key points about DF and suggest new directions for future research. First, similar to prior research, which shows

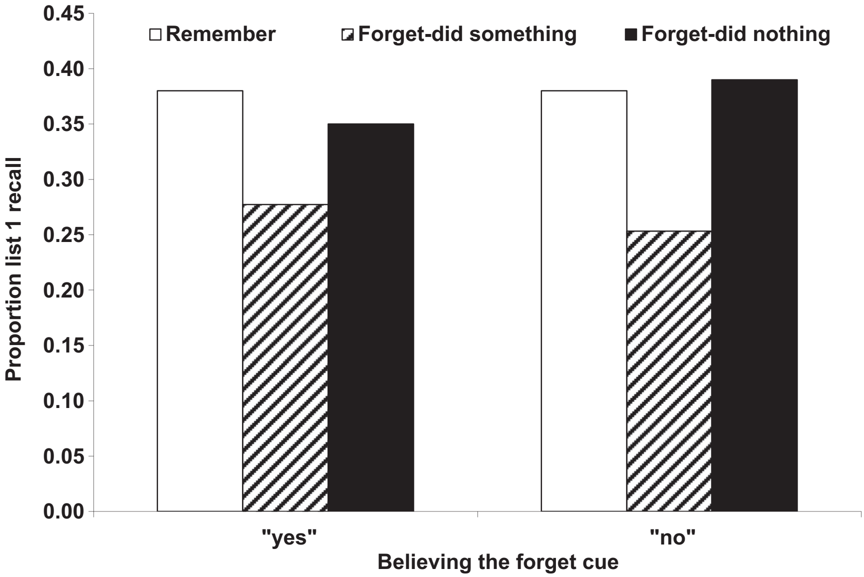

Figure 4.2 Directed forgetting impairment as a function of belief in the cue and controlled forgetting strategies. *(New figure based on the data reported in Foster and Sahakyan (2011)).*

that intention to learn, per se, is not important for remembering, but rather what people do to learn the information is what matters (Hyde & Jenkins, 1973; Mandler, 1967; Postman, 1964), the findings of Foster and Sahakyan (2011) demonstrate that merely believing the forget cue is not sufficient for demonstrating DF; what matters is what people do in response to the forget cue (see also Mulji & Bodner, 2010). Second, because believing the forget cue is not critical for obtaining DF, future research should be less concerned about staging an event to make the forget cue more believable (e.g. simulating a computer crash in the middle of the experiment), and should instead be more vigilant about reporting whether participants engaged in DF or not. For example, the absence of DF in certain conditions or certain populations could simply be linked to "doing nothing". Finally, understanding why some participants engage in DF whereas others do not should become a priority for future research as it will help to provide insights for understanding DF in special populations.

4.2. Why do Some People Fail to Engage in DF?

Are "do nothing" participants simply unmotivated to be in an experiment? If so, the remember group should recall much more than they do. In fact, though, the "do nothing" people show recall levels comparable to that of the remember group. Maybe some people simply cannot formulate a strategy in a brief period of time between the lists. Most experiments deliver the forget cue and move on to the second list right away. Indeed, some of the participants in our experiment reported "*I didn't do anything in particular, there wasn't enough time*," or "*I wanted to forget but didn't know how to forget, so I didn't do anything*," suggesting that extending the postcue time might increase the chances of obtaining DF or produce larger-than-normal effects. These reports also indicate that one may want to forget but be unable to formulate a strategy. It is essential to be mindful of these issues when interpreting DF performance among special populations, who may have difficulty self-initiating a controlled strategy.

Metacognitive beliefs could also play a role in why some participants fail to engage in DF. For example, if participants have a preexisting belief that their memory is not good even before they attempt to memorize the material in an experiment, then they may be less likely to engage in DF because they may believe that they already forgot. Indeed, a study in our lab with older adults revealed this problem (Sahakyan, Delaney, & Goodmon, 2008). Many older adults, in response to the forget cue, spontaneously volunteered that they had already forgotten. Not surprisingly, they did not engage in DF. We consequently modified DF instructions that emphasized the need to engage in forgetting despite the beliefs that they already forgot, and we found that such instructions were more conducive for obtaining DF with older adults than were the standard instructions.

4.3. Beliefs about Memory Developed during the Experiment

Participants could also develop beliefs about their memory *during* the experiment as they gain experience memorizing the materials. Such beliefs might also play a role in whether they engage in intentional forgetting. For example, if some participants do not feel confident about their ability to remember L1 after attempting to memorize it, they may see no need to engage in DF. Support for this hypothesis comes both from older adults (Sahakyan, Delaney, & Goodmon, 2008) and from a study involving mainly college adults (Sahakyan, Delaney, & Kelley, 2004). After presenting L1, we asked college participants to predict how many words from that list they would be able to recall during the final test (i.e. an aggregate judgment of learning (JOL)). Participants' predictions ranged from the minimum of 20% (equivalent to three words) all the way to 80% (equivalent to 12 words). After providing the JOLs, they received a forget or remember cue, followed by L2.

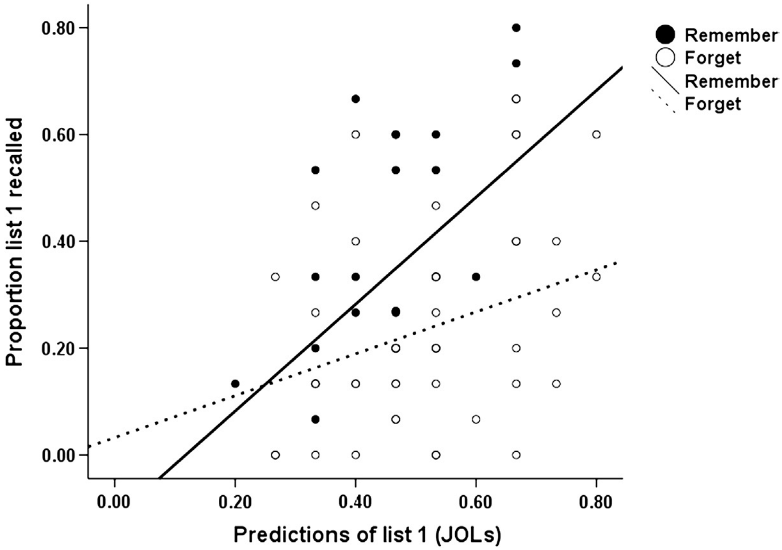

Figure 4.3 Directed forgetting impairment as a function of cue and aggregate predictions of list 1 learning. *(New figure based on the data reported in Sahakyan, Delaney, and Kelley (2004))*.

Figure 4.3 shows L1 recall during the final test as a function of JOLs and the cue. As the figure demonstrates, DF impairment was larger when participants anticipated recalling many L1 items than when they anticipated recalling few L1 items. Although we did not collect forgetting strategy reports in that study, it is possible that the more people expected to remember from the first list, the more likely they were to engage in DF. In contrast, if they felt they could not remember L1 that well, they may have been less likely to engage in DF. Individual differences in JOLs could be correlated with other variables that are involved in this effect (e.g. higher intelligence or working memory capacity), but these data are consistent with the idea that beliefs about memory could contribute to list-method DF (for a demonstration of how item-by-item JOLs affect item-method DF, see Foster & Sahakyan, 2012).

4.4. Preexisting Beliefs about Intentional Forgetting

In an unpublished study, Foster and Sahakyan investigated whether preexisting beliefs about whether it is possible to intentionally forget something might play a role in who engages in DF and who does not. People who have beliefs that it is impossible to forget something after it has been learned might be less likely to attempt DF. We asked participants who were participating in a departmental mass screening "*Do you think it is possible to make yourself forget something after you have learned it*?" The answer required a yes/no response. Out of 130 participants who responded to this question, 49% indicated "yes", and 51% indicated "no". These participants were later invited to take part in a DF experiment in our lab (they were unaware of the nature of the experiment). At the end of the experiment, they indicated what they did in response to the forget instruction. Their responses were coded as "doing something" or "doing nothing" (similar to Foster & Sahakyan, 2011). Replicating previous studies, 29% of the forget group reported "doing nothing". Interestingly, we found no correlation between peoples' beliefs about whether it is possible to intentionally forget and whether they engaged in DF ($\varphi = -0.04$). Most importantly, the magnitude of DF impairment was virtually identical among those who believed it is possible to intentionally forget and those who did not believe it is possible. What mattered for DF was whether participants engaged in deliberate strategies to forget, replicating previous work of Foster and Sahakyan (2011). The results are shown in Figure 4.4.

Overall, studies from our lab have consistently found that DF does not emerge among participants who report "doing nothing" in response to the

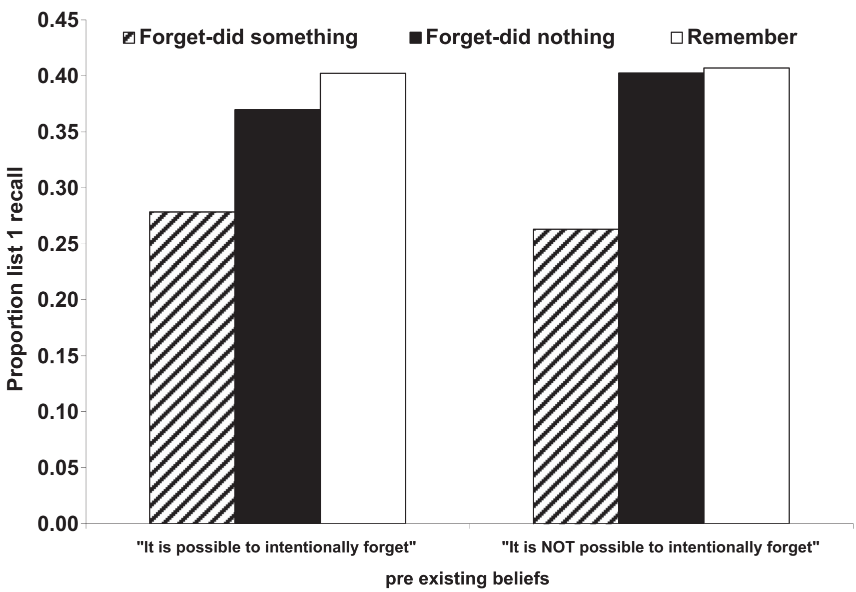

Figure 4.4 Directed forgetting impairment as a function of preexisting beliefs about intentional forgetting and controlled forgetting strategies. *(Based on unpublished data collected by Foster and Sahakyan.)*

forget cue. Engaging in DF does not appear to be linked to whether participants trusted the forget cue, or whether they think it is possible to forget things once they have been learned. A better understanding of why some healthy college adults do not engage in DF is warranted, because it could help explain DF performance in special populations. Failures to obtain DF with certain populations may not necessarily reflect an inability to inhibit. Instead, they may indicate that certain populations may select to do "nothing" in response to the forget cue, but the reasons for such decisions could be more complex and not necessarily linked to their ability to purposefully forget. They may be unable to formulate a strategy, or they may not see a need to engage in forgetting.

4.5. Forgetting Strategies and Their Relative Effectiveness

In recent years, Nathan Foster and Lili Sahakyan collected many strategy reports from participants in list-method DF studies. We did not analyze them in published papers due to smaller sample sizes. For this review, we combined several studies to compile a dataset of 290 participants (with 145 in the forget group and 145 in the remember group) to investigate whether DF depends on forgetting strategy or not. In each study, participants studied 15 unrelated nouns per list at 4 s/item, with the remember or forget cue between the lists, but no other break/delay. A short filler task followed L2, and then L1 was tested before L2. Then participants in the forget group were asked, "*What did you do to forget L1*? If you clearly remember what you did, please write that down. If you do not remember clearly, please do not make up anything at this point. If you did not do anything in particular to forget the words, or if you specifically tried to remember them, please indicate this."

The responses were grouped into six categories, and recall was assessed as a function of different strategies. The results are summarized in Figure 4.5. Reports of "stopping repeating L1 words," or "engaging in diverting thoughts" are more unambiguously interpretable compared to reports like "stopping thinking about L1 words," "clearing one's head," or "focusing on the upcoming list," because it is less clear how they went about those processes. Thus, interpreting the relative effectiveness of these strategies requires that we be mindful that some of these strategies may not be mutually exclusive.

It is nevertheless clear from Figure 4.5a that to obtain DF impairment, participants had to engage in some controlled behaviors; doing nothing in response to the forget cue did not produce DF impairment. Additionally,

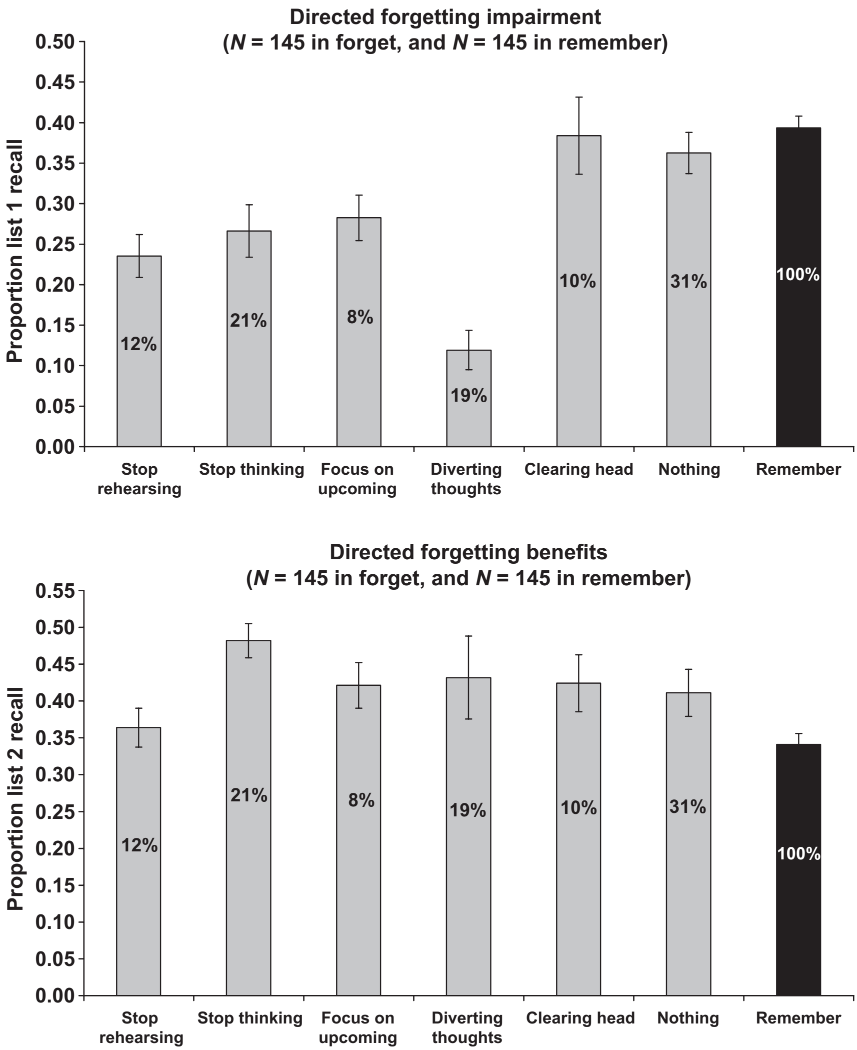

Figure 4.5 Directed forgetting costs (a) and benefits (b) as a function of forgetting strategies reported by forget-group participants in a dataset containing $N = 145$ forget-group participants and $N = 145$ remember-group participants. The black bar represents the remember condition, and the gray bars represent the subgroups of the forget condition broken down by reported strategy. The numbers inside the bars indicate the percentage of participants in each condition.

engaging in diversionary thoughts produced the biggest DF impairment compared to other strategies, some of which did not even produce DF (e.g. clearing one's head). Finally, despite the variability in the size of the DF impairment as a function of various strategies, nearly all strategies led to DF benefits, including "doing nothing" to forget (Figure 4.5b). We will revisit this issue in Section 7, where we specifically discuss the mechanisms producing DF benefits.

5. CONTEXT CHANGE AS AN EXPLANATION FOR DF IMPAIRMENT

Once forgetting is strategically initiated, we believe that it takes place because of changes in mental context—an explanation first proposed by Sahakyan and Kelley (2002). Our thinking was greatly influenced by formal models of episodic memory, according to which people encode both the *content* of the item during learning and various *contextual features* or attributes that are present in the background (e.g. Anderson & Bower, 1972; Estes, 1955; Gillund & Shiffrin, 1984; Hintzman, 1988; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988; Tulving, 1983). Contextual information includes environmental, spatial–temporal, emotional, and mental states in which the item is experienced. Context cues play a particularly important role in free recall because the initial retrieval attempt is solely guided by the context cues that participants use to search their memory (e.g. Howard & Kahana, 2002; Mensink & Raaijmakers, 1988). The success of retrieval is a function of the overlap between the cues used to search memory and the contents of memory. Recall suffers when there is a low overlap between the contextual features present at encoding and the contextual features present at test (e.g. Godden & Baddeley, 1975).

Based on these ideas, Sahakyan and Kelley (2002) proposed a mental context-change account of DF, according to which the forget instruction encourages participants to abandon the contextual cues that were prevalent during L1 encoding and to sample new contextual cues for L2 encoding, thereby segregating the two lists as separate events. The DF impairment arises because retrieval context better matches L2 than L1 encoding context, producing forgetting of L1 items in the forget group. The original context-change account further explained that DF benefits arise because of reduced proactive interference on L2 due to contextual differentiation (Sahakyan & Kelley, 2002). We will argue later that changes in encoding strategy produce a larger impact than the reduced proactive interference in most studies, but that the reduced proactive interference effect is likely a real effect that is difficult to assess through overall recall rates.

The rationale for proposing the mental context-change account was triggered by the reports of participants who indicated engaging in distracting thoughts in order to forget. Figure 4.5 confirms that this strategy leads to the largest DF impairment (although other strategies could also contribute to DF). We reasoned that engaging in diversionary thoughts could change peoples' mental context between L1 and L2. Recently, Lehman and Malmberg (2009) developed a computational model of DF within the framework of search of associative memory, by capitalizing on our notion of context change between the lists, and the difficulty reinstating L1 context during the test in the forget group.

In this section, we review the findings from various conditions that were predicted to interact with list-method DF based on the context-change mechanism along with other theoretically relevant findings.

5.1. Mental Context-Change Paradigm and Its Relation to DF

In the process of empirically testing the context-change account of DF, Sahakyan and Kelley invented what has come to be known as the *mental context-change/diversion paradigm*. This paradigm involves asking participants to engage in diversionary thoughts following L1 learning, and examining the impact of such activity on memory. Several years of investigation have demonstrated that mental context-change produces forgetting much like the explicit instructions to forget. A variety of tasks were shown to create forgetting, such as imagining being invisible (Sahakyan & Kelley, 2002), thinking about the childhood home (e.g. Sahakyan & Delaney, 2003), daydreaming about vacations (e.g. Delaney, Sahakyan, Kelley, & Zimmerman, 2010), or wiping a computer monitor (e.g. Mulji & Bodner, 2010). In contrast, other tasks do not lead to forgetting, such as solving arithmetic problems (Sahakyan & Delaney, 2003), counting forward or backward (Sahakyan, Delaney, & Goodmon, 2008), briefly chatting with an experimenter (Aslan & Bäuml, 2008), or waiting quietly (Sahakyan & Kelley, 2002). Why some tasks create mental context-change whereas others do not is not fully clear and requires future research.

Whenever context-change effects were found, they were found irrespective of encoding strategies (Sahakyan & Delaney, 2003), across individual differences in working memory capacity (Delaney & Sahakyan, 2007), across age-related differences (Aslan & Bäuml, 2008; Sahakyan, Delaney, & Goodmon, 2008), across serial position effects (Sahakyan & Foster, 2009), and the boundary conditions that determine whether DF is obtained, such as the need for L2 learning (Pastötter & Bäuml, 2007; but see Unsworth et al., 2012). Thus, the mental context-change paradigm produces many behavioral effects that are similar to DF. Interestingly, the retrieval dynamics in the mental context-change paradigm were shown to be similar to the environmental context-change paradigm across several dependent measures including recall latencies, serial position effects, first response functions, and the patterns of response transitions during recall captured through conditional response probability functions (Unsworth et al., 2012). Thus, the mental context-change paradigm produces behavioral effects similar to not only DF but also the environmental context-change paradigm.

Some recent findings, however, have revealed differences across DF and the mental context-change paradigm. For example, although Pastötter and Bäuml (2007) reported that L2 learning is critical for obtaining both DF and mental context-change effects, Unsworth et al. (2012) obtained significant forgetting using a mental context-change paradigm without involving L2 learning. Several methodological differences exist between these two studies, most notably the list length, which was much longer in the Unsworth et al. study (40 items) than is typically employed in the DF studies. The use of such long lists may affect the likelihood of reinstating the context of the items from the beginning of the list because those items were studied long ago. Future research should consider more rigorous investigation of the list-length factor, including examining whether long lists might also lead to a DF effect without the need for L2 learning.

5.2. Recognition Tests and Indirect Tests of Memory

The majority of list-method DF studies failed to obtain a DF effect on a standard recognition test (e.g. Basden et al., 1993; Bjork & Bjork, 2003; Block, 1971; Conway et al., 2000; Elmes, Adams, & Roediger, 1970; Geiselman et al., 1983; Gottlob & Golding, 2007; Smith, Barresi, Gross, & 1971; Racsmány, Conway, Garab, & Nagymáté, 2008; Reitman, Malin, Bjork, & Higman, 1973; Whetstone, Cross, & Whetstone, 1996; Zellner & Bäuml, 2006). Although two relatively recent studies obtained DF benefits on L2 using longer study lists than were previously employed (e.g. Benjamin, 2006; Sahakyan & Delaney, 2005), they too failed to obtain DF impairment. In contrast to the list-method, the item-method DF effect was always obtained on recognition tests (e.g. Basden et al., 1993; MacLeod, 1999). Implicit memory tests have also been shown to be insensitive to the list-method DF, including both word stem and fragment completion tests (Basden et al., 1993; Bjork & Bjork, 1996; but see Koppel & Storm, 2012), tests of primed word association or general knowledge (Basden & Basden, 1996, 1998), and lexical decision (Racsmány & Conway, 2006). In all these tasks, forget items show priming that is equivalent to remember items.

Because performance on the forget and the remember items did not differ on recognition tests (or on implicit tests), the inhibitory account proposed that those items were released from inhibition by virtue of being represented during the test (e.g. Bjork, 1989; Bjork & Bjork, 1996, 2003; Geiselman et al., 1983; Zellner & Bäuml, 2006). Although this explanation is still rather popular, we think of it more as a description rather than a specific assumption or an explanation offered by theory. The release of inhibition was proposed mainly to reconcile the discrepant findings across the list and the item methods. However, there is nothing *a priori* in the inhibitory account to propose a release from inhibition on certain tests. For example, assumptions of release are not made for the RIF phenomenon, which was discovered much later than DF, and which is usually observed on recognition and lexical decision tests (e.g. Aslan & Bäuml, 2011; Hicks & Starns, 2004; Soriano, Jiménez, Román, & Bajo, 2009; Spitzer & Bäuml, 2007, 2009; Veling & van Knippenberg, 2004; Verde, 2004). If list-method DF was observed on various memory tests just like the item-method, assumptions of release would not be made and the inhibitory interpretation of DF would probably not even be proposed.

We have interpreted DF from the perspective of our contextual account, and have a different take on the null effects of DF in various direct and indirect tests. For example, the absence of DF in implicit memory tests is consistent with the contextual account. Implicit tests do not require retrieval of contextual information because they do not direct attention to the original study episode. Since they do not require reinstating prior episodic context, we would not expect to observe DF on such tests. Environmental context effects also are not found on implicit tests (e.g. Parker, Gellatly, & Waterman, 1999).

The absence of DF in recognition is consistent with a body of literature that shows that environmental context effects typically are not detected in recognition (e.g. Godden & Baddeley, 1975). That said, the meta-analysis of that literature revealed that under certain conditions, environmental context effects do emerge in recognition, and those effects are larger when the encoding processes are primarily nonassociative in nature (for a review, see Smith & Vela, 2001). For example, a number of studies using nonwords or unfamiliar faces as stimuli detected environmental context effects in recognition (e.g. Dalton, 1993; Krafka & Penrod, 1985; Macken, 2002; Malpass & Devine, 1981; Russo, Ward, Geurts, & Scheres, 1999; Smith & Vela, 1992). Building on those findings, we predicted and obtained DF in recognition using nonword stimuli, despite not obtaining such effects using words (Sahakyan, Waldum, Benjamin, & Bickett, 2009).

In another experiment, we obtained DF in recognition using word stimuli by manipulating whether the test required distinguishing the targets from similar or dissimilar distractors (Sahakyan et al., 2009, Experiment 3). Similar distractors involved plurality-reversed versions of the target items, whereas dissimilar distractors were novel words. We hypothesized that plurality discrimination would engage more direct retrieval of contextual information during the test, enabling us to detect DF in recognition even with word stimuli. Indeed, the results confirmed our predictions: There was no DF in the dissimilar distractor condition, but there was DF in the similar distractor condition.

In the same year, Lehman and Malmberg (2009) reported DF in recognition of word stimuli using an exclusion/inclusion manipulation. They reasoned that typical recognition tests ask participants to endorse any item studied during the experiment, regardless of which list it appeared on (i.e. inclusion condition). Under such conditions, participants do not have to rely on context cues to differentiate one study episode from another, and they can rely on item familiarity as the basis of recognition. In contrast, in an exclusion condition, participants have to endorse only the words from a designated list (either L1 or L2), and to reject the words that come from a different list (or new words). To perform such a task accurately, participants have to rely more heavily on context cues to differentiate one list from another. Overall, the exclusion recognition is closer to what is needed to perform free recall than inclusion recognition is, and indeed Lehman and Malmberg (2009) predicted and obtained robust DF under exclusion instructions.

Plurality discrimination, exclusion recognition, and nonword recognition all rely on retrieval of contextual information, and hence all these conditions showed DF in recognition. Recently, Racsmány, Conway, Garab, and Nagymate (2008) used a remember/know procedure and also obtained DF in recognition, but only for "remember" responses. Although one might be tempted to conclude that DF impairs all types of contextual information, the results from associative recognition further refine the nature of contextual information that is important in DF. Specifically, Hanczakowski et al. (2012) did not obtain DF on an associative recognition test, despite it being a recollection-laden test. Although there was no effect in associative recognition, within the same study, DF affected list discrimination judgments. Therefore, the authors argued that DF does not lead to a generalized impairment in recollection of all contextual details. Instead, DF affects primarily global contextual information that differentiates L1 from L2, without necessarily impairing interitem associations between the words within the list. Finally, Gottlob and Golding (2007) demonstrated that despite the lack of list-method DF in the item recognition, the forget instruction impaired source memory for color and case of L1 items.

Overall, the reports of significant list-method DF in item recognition suggest that the typical conditions of recognition rather than the recognition test itself were the reasons for the observed lack of DF in previous research. When stimuli or test conditions encouraged greater utilization of global contextual information, DF was observed in recognition. These results challenge the inhibitory interpretation because it does not *a priori* predict the conditions under which the release can or cannot be obtained. In contrast, these results fit well with the context account. As the need to retrieve context increases, the likelihood of seeing list-method DF in item recognition or recognition of source also increases.

One of the earlier explanations of DF involved selective rehearsal, and some researchers lately acknowledged that this account should not be prematurely dismissed for intentional learning (e.g. Benjamin, 2006; Sheard & MacLeod, 2005). Considering that people often report stopping rehearsal in response to the forget cue and that engaging in this type of strategy produces DF impairment, we briefly discuss the results of recognition findings from the perspective of selective rehearsal. The absence of DF in recognition tests is particularly problematic for the selective rehearsal account because recognition and recall are sensitive to elaborative rehearsal (Geiselman & Bjork, 1980), and recognition is even more sensitive than recall to rote rehearsal (Benjamin & Bjork, 2000; Craik & Watkins, 1973). Thus, regardless of nature of the rehearsal process involved in DF, terminating rehearsal of to-be-forgotten items should have negative consequences for recognition memory. We note that recognition is typically not at ceiling despite the short lists used in many DF studies, and even longer lists do not lead to DF impairment in recognition (Benjamin, 2006; Sahakyan & Delaney, 2005). Thus, any explanation that attempts to revive the selective rehearsal account must successfully deal with this limitation, and should also be able to explain why recognition effects emerge under the specific conditions described above.

5.3. Mental Reinstatement of Context

The context-change account assumes that participants rely on the recent L2 context to probe their memory for L1, and because that context mismatches the L1 context in the forget group, it leads to impaired access to L1 items. The context account predicts that if participants could reinstate L1 context at the time of test, this should reduce DF.

The environmental context-change studies imply that participants do not automatically reinstate the original environmental context when they are tested in a new context, otherwise we would not be observing forgetting (e.g. Godden & Baddeley, 1975). Either participants do not attempt contextual reinstatement in those studies, or they attempt to reinstate the original context but are unsuccessful at it. We suspected that reinstatement of mental context (i.e. internally generated context) might be even more challenging than reinstatement of environmental context. To make L1 context more memorable and distinctive, we played the music from Star Wars at the start of the experiment. Before final test, Sahakyan and Kelley (2002) asked participants to mentally reinstate the episodic context that was prevalent during L1 encoding by guiding them through similar types of steps as previously employed in the environmental context literature (e.g. Smith, 1979). Reinstatement of L1 context did not eliminate the DF effect. However, it significantly reduced it both in the mental context-change condition and in the forget group, suggesting that the interaction between the contents of memory and the contextual cues used during the test is a critical ingredient for obtaining DF.

If L1 items were inhibited as a result of DF, then why did they recover from contextual reinstatement not just in the context-change group but also in the *forget* group? To explain the recovery, the inhibitory account would have to assume that context plays an important role in retrieval, and that changing or reinstating the context at the time of final test has consequences for memory. Because these are general claims made by many formal theories of memory, it is unclear what is gained by postulating an inhibitory mechanism if it must still explain the results by resorting to the contextual mechanisms.

5.4. Cuing with Context Cues vs Category Cues

Recent research manipulated the structure of the study list combined with specific ways of cuing memory at the time of test, and predicted the factors that should eliminate DF vs preserve it (Lehman & Malmberg, 2011a). Participants studied two lists. L2 always contained unrelated items, but L1 consisted of unrelated items for some participants vs items drawn from a single category for others. On the test, memory was probed either with a context cue (e.g. "retrieve the items from L1") or using the category cue (e.g. "use the word *clothing* to help you retrieve the words from L1"). The authors predicted that if DF impairment is due to context change, then it should be reduced with category cues, which will reinstate the specific L1 context more than will more general temporal context cues (e.g. Raaijmakers & Shiffrin, 1981). Successful initial recall would in turn provide a good retrieval cue for subsequent items. Replicating prior work, DF was obtained with unrelated lists. There was also significant DF with the categorized list as long as the test cue was just the temporal context. In contrast, when a category cue was used to test memory, the costs were eliminated, as predicted. Lehman and Malmberg's (2011a) findings highlight the importance of the interaction between retrieval cues and the contents of memory at test: The same categorical list can either show DF or be released from DF depending on the type of retrieval cues utilized during the test. The more specific the cue is to the L1 study context, the less likely DF is to be observed.

Note that another study conducted by Wilson, Kipp, and Chapman (2003) obtained a different outcome with categorical lists, arguing that categorical lists were resistant to DF. However, before the study of each list, Wilson et al. (2003) drew special attention to the fact that the lists in their study were either unrelated or categorical in nature. Their test instructions were not described in sufficient detail to infer what type of retrieval cues were given to participants. Thus, methodological differences between the two studies may be responsible for the discrepant findings. Note that Sahakyan (2004) also obtained intact DF using different categorical items across the lists while probing memory with temporal context cues.

5.5. Encoding Strength: Item vs Context Strength

Does deeper/elaborate encoding of material protect against DF? Sahakyan and Delaney (2003) had participants encode both lists either via the story mnemonic or via rote rehearsal. Despite robust differences in overall memory performance, they obtained equivalent DF impairment in both encoding conditions, suggesting that interactive encoding does not protect against DF.

Our subsequent research suggests that certain ways of strengthening items make them more susceptible to DF than other ways of strengthening (Sahakyan, Delaney, & Waldum, 2008). Relying on research by Malmberg and Shiffrin (2005), who suggested a more complex view of encoding strength by differentiating it into an item strength and a context strength, we derived and tested predictions of the contextual account of DF. Specifically, Malmberg and Shiffrin (2005) suggested that certain strengthening manipulations increase the strength of the items without affecting their associations with the list-context (e.g. depth of processing or extrastudy time), whereas other manipulations lead to strengthening of both item and contextual information in the memory trace (e.g. spaced presentations).

Based on their findings, we predicted that spaced items should suffer more when context changes because item-to-context associations are stronger for spaced items. Indeed, our study showed that spaced items showed larger DF impairment than did massed items. Similar results have also been obtained in an environmental context-change study (Isarida & Morii, 1986). In contrast to the spacing manipulation, we predicted that the depth of processing or the extrastudy time manipulations should not interact with DF, and indeed the results confirmed our predictions. Even though items studied for longer duration were remembered better than items studied for shorter duration, they showed equivalent impairment from DF. Taken together, these results suggest that what determines interactions with DF is not the strength of the items per se, but rather the strength of item-to-context associations. Along the same lines, Waldum and Sahakyan (2012) found that statements incongruent with one's attitude show greater DF impairment than do congruent statements. Using a variety of memory paradigms, they confirmed that greater DF of incongruent statements was driven by stronger associations with episodic context, entirely consistent with the context account.

5.6. Contextual Variability

In an unpublished study, Sahakyan and Foster crossed DF with manipulation of context variability (CV), which is the number of different semantic contexts in which a given item occurs (e.g. Dennis & Humphreys, 2001; Steyvers & Malmberg, 2003). The rationale for their study was that although low-CV items have a memory advantage in both recognition and free recall, they tend to suffer more from environmental context manipulations between study and test (Marsh, Meeks, Hicks, Cook, & Clark-Foos, 2006). Marsh et al. (2006) argued that during encoding, low-CV items form stronger item-to-context associations compared to high-CV items due to a smaller fan of preexisting contextual associations. Therefore, when context changes between the study and the test, low-CV items suffer more than high-CV items do.

Consequently, we expected that DF should also affect low-CV items more than high-CV items. We tested 46 participants randomly assigned to forget or remember groups. Each participant studied two lists of 14 items selected from Steyvers and Malmberg's (2003) appendix. Half of the items within each list were low-CV and the other half was high-CV words. Because CV correlates with word frequency (Steyvers & Malmberg, 2003), we controlled word frequency by selecting the items from the low-frequency category similar to what was done by Marsh et al. (2006). During

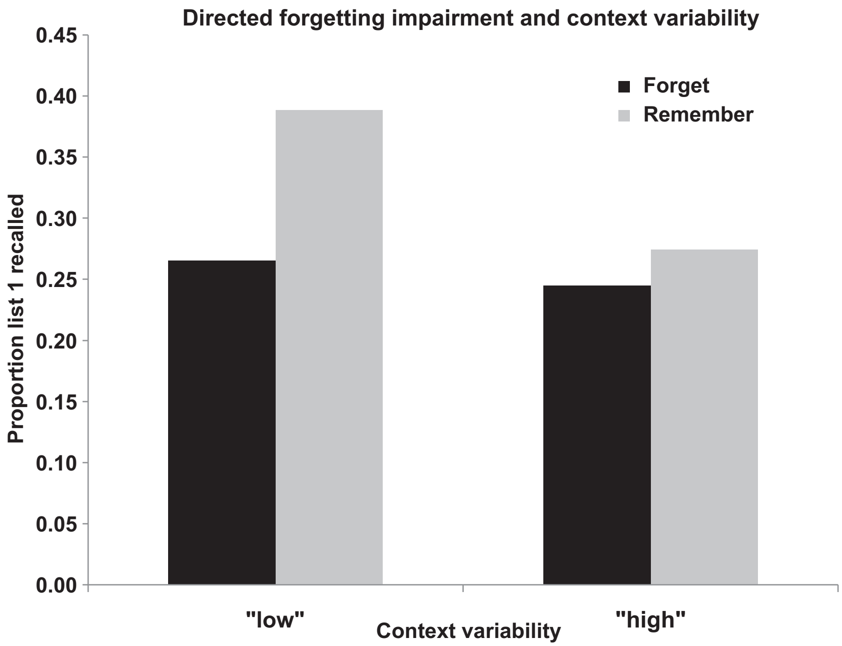

Figure 4.6 Directed forgetting impairment as a function of context variability of studied items. *(Based on unpublished data.)*

the test, L1 was always tested before L2. Figure 4.6 summarizes the findings. DF impairment was much greater for low-CV items than for high-CV items, consistent with the predictions of the context account.

5.7. Extralist Cued Recall

Most list-method DF studies employ free recall tasks partly because recognition and indirect memory tests are insensitive to DF. Sahakyan and Goodmon (2010) used a different testing procedure that allowed them to contrast the predictions of the inhibitory and context accounts. Specifically, participants studied two lists of items (called targets). The test provided cues that were meaningfully related to the targets, but were never studied with them (hence the name extralist cued recall). For example, the studied word *planet* could be tested with the cue *universe*. Extralist cued recall is affected both by target characteristics and by cue characteristics (for reviews, see Nelson, McKinney, Gee, & Janczura, 1998; Nelson & McEvoy, 2005), and it provides an excellent opportunity to test the opposite predictions of the inhibitory account and the context account within the same study.

Sahakyan and Goodmon (2010) conducted five experiments manipulating different sources of implicit associative strength. Crucially, some of their experiments varied the implicit strength of the targets while holding the cue strength constant, whereas other experiments held the implicit strength of targets constant while varying the strength levels of the test cues. Across all experiments, they found greater DF from the strong conditions compared to the weak conditions, regardless of whether strength referred to the targets or the test cues. This pattern of findings is exactly what is predicted by the contextual account: Presentation of the extralist cue should not diminish the importance of recovering contextual information during the test. Reinstating the context helps to differentiate the studied item from other items implicitly activated by the test cue. According to the PIER model which explains extralist cued recall (Nelson, Goodmon, & Ceo, 2007), context cues and extralist cues combine together to elicit recall of the target. Thus, the context account makes a strong prediction that there should be greater effects of DF in the strong than weak conditions regardless of whether the study varied the target strength level or the cue strength level.

In contrast, the inhibitory view assumes that items can be released from inhibition with provision of certain cues (e.g. copy cues of the items on the recognition test). Hence, the inhibition account would predict that stronger test cues would be more successful at releasing the targets from inhibition than weaker cues, and hence DF should be smaller not greater in the conditions where cue strength is high. The results did not support this prediction. The findings of Sahakyan and Goodmon (2010) suggest that items/targets were not inhibited at the level of individual item representation because there is no way for the cognitive system to determine which targets should be inhibited more and which targets should be inhibited less until a test cue is provided. The effects in their study were instead entirely driven by the combination of test cues and context cues, consistent with predictions of the context account.

5.8. Importance of L2 Learning

A number of studies revealed boundary conditions that are needed to observe DF impairment. For example, replicating prior work by Gelfand and Bjork (1985; cited in Bjork, 1989), Pastötter and Bäuml (2007) showed that in the absence of L2 learning, one does not observe DF impairment. They later showed that DF is reduced with decreasing length of L2 (Pastötter & Bäuml, 2010). Finally, Conway et al. (2000) as well as Macrae, Bodenhausen, Milne, and Ford (1997) demonstrated that divided attention during L2 learning by a concurrent task reduces DF impairment.

According to the inhibitory view, in the absence of L2 learning, there is no need to inhibit L1 items because inhibition is an adaptive mechanism invoked to reduce interference (e.g. Conway et al., 2000; Barnier et al., 2007). When there is no L2 learning, there is no need to invoke inhibition. The inhibitory view thus assumes that L1 items intrude during L2 learning and hence must be inhibited (e.g. Conway & Fthenaki, 2003; Racsmány & Conway, 2006). However, in the domain of RIF, it is generally assumed that simply studying a list of items should not trigger inhibitory processes; inhibition should be triggered only in response to competitive retrieval (e.g. Anderson, 2003; Storm & Levy, 2012). It is not fully clear why the mere study of L2 would trigger competitive retrieval of L1 items.

According to the context-change account, DF arises not simply because of context change between the lists, but because participants experience difficulty reinstating the context of the previous list. In the absence of L2 learning, reinstating the context of the only list that was studied is not an issue because it was the only list that was studied (e.g. Jang & Huber, 2008). Thus, according to the context-change view, one would not expect to observe DF. In contrast, in the presence of L2, participants have to rely on context cues to distinguish between the two lists and reinstate the context of the previous list. Also, in the environmental context literature, the effects of the first-order paradigms, where a single list is studied and tested either in the same or a different context, are much smaller compared to the second-order paradigms that involve multiple lists (e.g. Eich, 1985; Fernandez & Glenberg, 1985; for a review of first- and second-order paradigms, see Bjork & Richardson-Klavehn, 1989). We interpret the impact of L2 length from a similar perspective. Enhanced length of L2 contributes to greater distancing of L1 context from the time of test and, as it gets more distant, the reinstatement of that context would also get more difficult, explaining why shorter L2 reduces DF. Finally, we interpret the results of divided attention as reflecting attentional demands of context encoding and context change. We expand on this view in the section discussing working memory interactions.

5.9. Related Items Across Lists

Several studies have shown that when items are related across the lists, DF impairment is eliminated (e.g. Barnier et al., 2007, Experiment 5; Conway et al., 2000; Sahakyan & Goodmon, 2007). These findings are not consistent with an inhibitory viewpoint that assumes suppression at the level of items. Specifically, one would assume that when the two lists involve related items, then L1 items may be more likely to come to mind during L2 learning and require greater inhibition compared to when the two lists are unrelated. However, the results show that related items across the lists *reduce* rather than increase DF.

Although items across the lists may be related to each other in a number of ways, our research showed that when L2 items remind participants of L1 items due to backward associations (e.g. *chip← chisel*), DF is eliminated, but when L1 items are related to L2 items via forward associations (e.g. *chip→ wood*), DF remains intact (Sahakyan & Goodmon, 2007; for arelated result using the item procedure, see Golding, Long, & MacLeod, 1994). We proposed several mechanisms by which related items could reduce DF. First, reminding could reinstate the context of L1 during L2 learning, thereby preventing or reducing contextual differentiation between the lists. Second, reminding could initiate retrieval of L1 items during L2 learning, thereby linking L1 items to the context of both lists. This would enhance the probability of their retrieval even if L2 context is used as the main retrieval cue during the test. Finally, reminding could strengthen L1 items via retrieval during L2, and enhanced strength of L1 items could also make them resistant to DF.

5.10. Retrieving L2 before L1

The meta-analysis suggests that output order does *not* affect the magnitude of DF impairment (Pastötter et al., 2012). In other words, whether L1 is tested first in the recall sequence, or whether it is tested after L2, the magnitude of DF impairment remains invariant. This is inconsistent with the inhibitory explanation because if L2 is tested first, then according to the inhibitory account, L1 items should be inhibited to allow retrieval of L2. Thus, the DF impairment should be larger when L1 is tested after L2 than when it is tested first in the recall sequence. However, Pastötter et al. (2012) showed that retrieving L2 first does not increase the magnitude of forgetting.

An even more extreme manipulation was employed by Basden, Basden, and Morales (2003), who gave multiple retrieval trials on L2 before the final test. Despite multiple retrieval trials on L2, Basden, Basden, and Morales (2003) and Basden, Basden, and Wright (2003) did not observe negative effects of such retrieval on L1 memory, even when their lists consisted of categorical items from the same taxonomy. Overall, these results suggest that retrieving L2 before L1 has no detrimental effects on DF impairment, contrary to what would be predicted by the inhibitory account.

5.11. Delay Effects

There is a scarcity of published studies that have examined DF after a delay. If DF impairment arises from the mismatch of the study and test contexts, then the impairment should dissipate with the passage of time because with delay, context no longer would favor L2 context over L1, and thus one would expect spontaneous recovery of L1 (e.g. Mensink & Raaijmakers, 1988). Similar predictions would be made also by the inhibitory account because items should not remain in an inhibited state indefinitely. We found only one published study that employed a delay manipulation with a full-DF design (Shapiro, Lindsey, & Krishan, 2006). The delay involved 15 min, and participants were tested either immediately, or after delay (but never in both). The immediate test revealed the costs and the benefits of DF. In contrast, delay eliminated DF costs, but did not affect the benefits, which remained preserved. Identical results were reported by Liu Xun in his unpublished dissertation (2001). The presence of significant L2 benefits is consistent with the encoding-based interpretation of DF benefits (Pastötter & Bäuml, 2010; Sahakyan & Delaney, 2005) because stronger encoding of L2 items should be preserved relatively in delay (e.g. MacLeod, 1975). Thus, the dissociating effect of delay is consistent with the dual-factor account of DF.

Two other delay studies were briefly discussed in two chapters (Basden & Basden, 1998; MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003), and they reported significant DF in a delayed condition. However, neither study included a remember control group, and DF was inferred from the R–F measure. Given that delay preserves the benefits (Liu, 2001; Shapiro, Lindsey, & Krishnan, 2006), the difference between the two lists would probably still be significant after a delay. Most important, the brief description of the methods in both studies makes it difficult to know whether the same participants took part in both the immediate and the delayed test, or whether delay was varied strictly between subjects. This could be a critical detail given the effects of testing on retention (e.g. Roediger & Karpicke, 2006). If the same participants participated in both conditions, then the immediate test could preserve DF on the delayed test. In other words, significant DF on the delayed test in some studies may be due to a testing effect. More research is needed to fully investigate delay effects in DF.

5.12. Individual Differences in DF as a Function of Working Memory Capacity

Another feature of the contextual change account is that attentional control has been theoretically linked to effective use of context (e.g. Unsworth & Engle, 2008). Therefore, one might expect that people with higher working memory capacity, which is an index of attentional control, would show larger effects of the forget cue and of the context-change task, as they would be better able to manipulate context. Indeed, this appears to be the case in several recent studies. Delaney and Sahakyan (2007) showed that in a typical two-list design, both L1 and L2 memory were positively correlated with working memory capacity. (This effect disappeared when everyone was instructed to make up a story using all the words on the list). However, working memory capacity was negatively correlated with L1 memory in the forget and context-change conditions in two experiments. In other words, the DF and context-change manipulations produced substantially more forgetting for high-span than for low-span participants. Consistent with the context-change account, both the context-change task and the forget task interacted with working memory capacity in the same manner. This basic result has been replicated with adults and children (Aslan, Zellner, & Bäuml, 2010), and in some reports, the benefits of DF were also linked to working memory capacity (Soriano & Bajo, 2007).

The effects of dividing attention during L2 (e.g. Conway et al., 2000) are consistent with the notion that context encoding requires attentional resources. Dual task could interfere with context encoding during L2, thereby giving contextual advantage to L1 compared to L2. It could also make the two lists sufficiently differentiated even in the remember group. Note that Conway et al. (2000) did not include a condition where attention was divided during both lists. Thus their attentional manipulation was confounded with the list.

5.13. Reexposure or Retrieval of L1 Items Affects Subsequent Memory

Goernert and Larson (1994) examined how part-set cuing affects list-method DF and found that when participants were provided with a subset of L1 items on the final test to use as cues for retrieval, their recall of remaining L1 items showed improvement in the forget group, whereas it suffered in the remember group compared to the noncued condition. These effects were replicated and extended by Bäuml and Samenieh

(2010, 2012a, 2012b), who showed that both part-set cuing and also selective retrieval of a subset of L1 items can enhance recall of the remaining L1 items in the forget group, while the same manipulations impaired recall in the remember group. Similar effects were obtained in the mental context-change paradigm (Bäuml & Samenieh, 2012a, 2012b). Prior studies also showed that reexposure to L1 items on an intervening test can release subsequent DF depending upon the nature of the intervening test. The typical design of these studies involves (1) giving an initial test, (2) reexposing all or a subset of items from both lists on an intervening test, and (3) administering a final recall test. The intervening tests included recognition (Basden et al., 1993; Basden et al., 2003; Bjork & Bjork, 1996), word fragment completion (Basden et al., 2003; Bjork & Bjork, 1996), implicit free association test (Basden et al., 1993), or a lexical decision test intermixing all studied items with unstudied items and nonwords (Racsmány & Conway, 2006). Although initial recall revealed significant DF in these studies, the intervening tests did not (neither recognition, nor lexical decision, nor fragment completion showed evidence of DF). Importantly, the final recall findings differed depending upon the nature of the intervening task. When the intervening test involved a recognition test that included a subset of L1 items as lures, DF was released on a final recall test. In contrast, intervening word fragment completion or lexical decision tests did not release DF.

We interpret these findings to suggest that retrieval of contextual information associated with the study episode is critical to reinstating access to L1 items (similar arguments have been made by Bauml et al. for the part-set cuing and part-set retrieval effects in the forget group). The fragment completion test is an indirect memory test because it does not direct the subject to go back to the original learning episode, and it is considered mainly data driven. Likewise, the lexical decision task does not require reinstating prior episodic context. The intervening recognition test, however, is a direct measure of memory because the subject is instructed to think back to the learning episode and make a decision about a specific item (note that Bjork & Bjork, 1996 included an exclusion instruction on the intervening test of recognition, which would further facilitate retrieval of contextual information). Thus, DF on the final free recall test is reduced when participants have to think back to the original study episode and reinstate its episodic context. In contrast, DF is unaffected by indirect tests that do not make any connection to the earlier study episode.

5.14. Summary

The evidence that contextual change underlies DF impairment comes from a wide range of experimental sources. The idea came about because participants in DF studies self-reported thinking of something else, and subsequent research confirmed that this strategy was effective in producing forgetting. Furthermore, context is used to explain many other findings in memory, so it is not invented specifically to explain DF, and can be modeled in context-based formal models of memory. Among the findings that are predictable from a context viewpoint are that recognition does not show DF unless it relies on retrieval of contextual information; that mentally reinstating the original study context releases DF; that indirect memory tests do not show DF because they do not redirect back to the context of the study episode; that items more strongly associated with episodic context show greater DF; that cuing the items with stronger extralist cues leads to larger rather than smaller DF; that having strong reminders on L2 can reduce DF; that having no L2 eliminates DF; that categorized items can be susceptible to DF or not depending on whether the test cue involves context cue or a category cue. In our view, the account is strongly supported by the existing data, and any future theory should strive to not only explain the effects that the context-change account predicts but also make novel predictions that cannot be handled by the current theory.

Shortly after we proposed the context-change account of DF, Anderson (2005) proposed that the context-change mechanism is compatible with an inhibitory interpretation of DF if inhibition is assumed to be the mechanism by which context change takes place. According to Anderson's interpretation, inhibition is a flexible mechanism that can be targeted at different levels of representation, including at a context representation. We are reluctant to embrace this notion because we do not see it as advancing our theoretical understanding of DF. It redescribes context change with new terms (i.e. inhibition of context) without specifying a new theory of inhibition of context. There is currently no evidence to suggest that context is inhibited as opposed to being changed across the lists in DF. Impaired memory of L1 items does not constitute any evidence of inhibition of context or indeed any inhibition at all (MacLeod et al., 2003). Equating mental context change with inhibition of context implies that any time we change what we are thinking about, we engage in inhibition. By this view, everything becomes inhibition, and thus impossible to know what does *not* involve inhibition. Inhibition of context also complicates the explanation of

DF because it implies two mechanisms, not one. If we assume that context serves as retrieval cue, then when one changes from context A to context B, context match becomes poorer for the previous context A. Adding inhibition would be imposing a second mechanism that assumes that changing from context A to context B requires inhibition of context A. To the best of our knowledge, no one has offered an inhibitory theory of environmental context-dependent memory (e.g. Godden & Baddeley, 1975). The impaired memory arising from the context change is instead explained in terms of the mismatch of cues being used to search memory and the contents of memory, which leads to reduced accessibility.

6. AREAS OF DISAGREEMENT ACROSS STUDIES

While reviewing the empirical findings, two broad areas of inconsistencies emerged as in need of further investigation and more careful analyses. These include L1 serial position findings, and the source misattribution and intrusion error findings.

6.1. Serial Position Effects

We identified five published DF studies that examined serial position functions in DF (Geiselman et al., 1983; Lehman & Malmberg, 2009; Pastötter & Bäuml, 2010; Pastötter et al., 2012; Sahakyan & Foster, 2009). Some of these studies used serial positions as a tool for testing various DF theories, whereas other studies plotted those functions without formally analyzing them. Although the studies agree that the effect of the forget cue on L2 is mostly evident at the start–middle of the L2 serial position curve, their results regarding L1 functions differed to some extent. For example, studies that analyzed serial position functions did not obtain greater DF costs from the primacy region of L1 curve (e.g. Pastötter & Bäuml, 2010; Pastötter et al., 2012; Sahakyan & Foster, 2009), whereas studies that reported serial position curves without formal analyses (or deviation measures) obtained greater DF from the primacy regions of L1 curves (e.g. Geiselman et al., 1983; Lehman & Malmberg, 2009). Lehman and Malmberg (2009) argued that context is more prominent at the start of the list, which could explain greater forgetting from the primacy region. A study reported by Sheard and MacLeod (2005) also obtained greater DF costs from L1 primacy region, although their study failed to obtain overall DF impairment. Thus, some studies obtain greater DF impairment in the primacy region whereas others do not. Note that Sahakyan and Foster (2009) obtained equivalent DF impairment across L1 serial position curves of performed actions phrases which typically do not produce primacy effects, suggesting that the magnitude of DF impairment may not be linked to the primacy effects. Considering these controversies and the methodological differences that could have given rise to these findings, more careful attention is needed to serial position effects in DF research.

6.2. List Differentiation in DF: Source vs Intrusion Errors

Some DF studies found that the forget cue impaired source discrimination judgments in recognition (e.g. Bjork & Bjork, 2003; Gottlob & Golding, 2007; Hanczakowski et al., 2012), whereas we explained these findings in terms of a bias to attribute L1 items to L2 during the recognition test (e.g. Sahakyan & Delaney, 2005). Regardless of whether this is a bias effect or a true source confusion effect, these findings overall suggest that DF impairs list differentiation. Somewhat different conclusions are reached when DF is examined via intrusion errors during free recall or list categorization errors during free recall (e.g. Geiselman et al., 1983; Golding & Gottlob, 2005; Lehman & Malmberg, 2009; Sahakyan & Delaney, 2010; Spillers & Unsworth, 2011). Some of these studies have found enhanced list differentiation as evidenced by reduced intrusion errors, whereas others have found the opposite. Although intrusion errors are notoriously difficult to study, they could nevertheless be informative because they could indicate the type of retrieval cues that participants rely on during free recall. For example, if people make many intrusions from a wrong list, this might suggest problems with the sampling stage, indication that they use "wrong retrieval cues" that better match one list than the other list (e.g. Spillers & Unsworth, 2011).

As list differentiation in DF is studied further, it is important to keep in mind that source errors and intrusion errors may not necessarily lead to the same conclusions because they represent opposite sides of the same coin. Source errors refer to the ability to retrieve the context given the item, whereas intrusion errors are more diagnostic of the cuing property of context, referring to the ability to retrieve the item given the context cue. If item-to-context and context-to-item associations are not necessarily equivalent in strength, or if ability to retrieve the context given the item as opposed to retrieving the item given the context were not identical, then the findings from source errors and intrusion errors may not necessarily produce the same empirical outcomes.

7. STRATEGY CHANGE EXPLAINS DF BENEFITS

Sahakyan and Delaney (2003) were the first to propose that the costs and benefits of DF might be dissociable, and that they might be explained by different mechanisms. A critical piece of evidence for their proposal was that it is possible to dissociate the costs and benefits (e.g. Benjamin, 2006; Conway et al., 2000; Joslyn & Oakes, 2005; Pastötter et al., 2012; Sahakyan & Delaney, 2005, 2010; Shapiro et al., 2006; Spillers & Unsworth, 2011; Zellner & Bäuml, 2006). How could that be true if a single process accounted for both results? While not all researchers agree what the two processes are, most now agree that separate mechanisms are needed to explain costs and benefits.

7.1. The Advent of the Dual-Process Account

The Sahakyan and Delaney (2003) dual-process account explains the benefits by the strategy-change mechanism. According to this explanation, the benefits of DF emerge because the forget cue encourages participants in the forget group to adopt a more elaborate encoding strategy for L2 items compared to those in the remember group. Thus, L2 benefits reflect a strategic encoding enhancement in the forget group.

We initially proposed this explanation because we evaluated participants' verbal reports of study strategies from Sahakyan and Kelley's (2002) study, which revealed that many participants adopted a better study strategy on L2 than L1, and the forget group was almost twice as likely as the remember group to switch encoding strategies (38% vs 22% in the Sahakyan and Kelley study). Therefore, in Sahakyan and Delaney (2003), we first reanalyzed Sahakyan and Kelley's (2002) data by statistically controlling the strategies, and we found that the effect of cue (F vs R) was much smaller compared to the effect of strategy switch, which accounted for the majority of variance in L2 recall. We also conducted two experiments that experimentally controlled the encoding strategy, with the same outcome. In Experiment 1, participants studied both lists using the same encoding strategy, and we did not obtain the benefits because everybody suffered massive proactive interference. In Experiment 2, everybody was told to switch study strategy between L1 and L2 (including the remember group), and everybody benefited from such switch and escaped proactive interference. In both experiments, we always obtained DF costs.

This is important to emphasize because we often hear from our colleagues that a switch of study strategies between the two lists may be part of mental context change and could explain the DF impairment. Although strategy switch could contribute to mental context change, it cannot explain why the DF impairment was still obtained in Experiment 2, when everybody switched strategies, including the remember group. According to this reasoning, there should be no DF costs if *everybody* changes mental context due to strategy shift. However, the costs are present regardless of whether people retain the same strategy on both lists, or change the strategy across the lists, which led us to propose the dual-process account. Studies investigating neural bases of DF also provide support for the two-factor account, and suggest different neural origins for the costs and the benefits of DF and mental context change (Bäuml, Hanslmayr, Pastotter, & Klimesch, 2008; Hanslmayr et al., 2012; Pastötter, Bäuml, & Hanslmayr, 2008).

In general, whenever we controlled encoding strategy in subsequent studies, we never obtained the benefits (e.g. Sahakyan & Delaney, 2005, 2010; Sahakyan, Delaney, & Waldum, 2008). Along the same lines, when participants performed simple action phrases during encoding (e.g. *break a toothpick*), we did not obtain the benefits, but the same action phrases encoded verbally without controlling for study strategy produced the benefits (e.g. Sahakyan & Foster, 2009). Although we explain L2 benefits in terms of encoding strategy switch, a related and similar account was proposed by Pastötter and Bäuml (2010), who argued that the benefits arise because the forget cue resets the encoding processes at the start of L2 learning. This proposal is analogous to the earlier selective rehearsal account, with an exception that it reserves selective rehearsal mechanism strictly for explaining L2 benefits (as opposed to the costs and the benefits as was initially assumed by the selective rehearsal). The primary reason for proposing a reset of encoding (as opposed to a strategy change) is that L2 serial position curves typically show much larger DF benefits in the primacy regions of the curve than elsewhere in the list (e.g. Geiselman et al., 1983; Lehman & Malmberg, 2009; Pastötter & Bäuml, 2010; Sahakyan & Foster, 2009). Pastötter and Bäuml (2010) argued that a switch to a better encoding strategy should be evident throughout the entire list, not just at the beginning of L2. This interpretation of strategy change is reasonable, but might be a bit extreme. Although many participants spontaneously report doing something different to encode L2 (or in addition to what they were doing to encode L1), it is quite possible that they do not persist with that strategy throughout the entire L2 learning. For instance, often participants report starting out with some strategy on L2, but further down the list they report being overwhelmed and being unable to keep it up. This could explain why the benefits of DF are more prominent in the early sections of L2 serial position curves than in later sections.

7.2. A Large-Scale Replication Study

Our dataset of $N = 290$ participants (Figure 4.5) provides reports of strategy in response to the forget cue, and also contains reports of participants' encoding strategy for both lists. We therefore reevaluated the strategy-change account in our dataset with a substantially larger sample. As Figure 4.5 demonstrates, despite using a variety of forgetting strategies with some being more effective than others in leading to DF impairment, nearly all strategies were associated with almost invariant magnitude of DF benefits. Interestingly, even the group that reported doing nothing to forget also showed the benefits. The "doing-nothing" group essentially shows a dissociation, whereby they show no DF impairment compared to the remember group, but somehow they show the benefits on L2 despite no forgetting of L1. If the benefits are driven by strategy change, then it is quite possible that these participants changed study strategy between the lists (perhaps in hopes that doing so might help them forget L1). We coded participants' encoding strategies for both lists, and created a variable that captured whether participants employed a more efficient strategy on L2 compared to L1. For example, any mention of doing something in addition to what they were doing during L1 encoding was coded as a *strategy switch* (e.g. rehearsing on L1, and rehearsing plus also grouping words together on L2).

Confirming previous work from our lab, we found that 28% of the forget group switched to a better study strategy on L2 compared to 10% of the remember group, X^2 $(1, N = 290) = 14.43$, $p < 0.001$. Although the rates of strategy switch were much lower in the University of North Carolina at Greensboro sample than in the previous samples of Florida State University and University of Florida participants, the basic pattern was similar—the forget group was more likely than the remember group to switch to better encoding on L2. Confirming the results of Foster and Sahakyan (2011), we also found that within the forget group, there was no relation between the likelihood of switching to a better study strategy on L2 and the likelihood of engaging in the forgetting strategy such as doing something vs nothing to forget L1 ($\varphi = -0.03$, $p = 0.73$). We therefore collapsed across that factor in the forget group, and analyzed L2 recall using cue (forget vs remember) and strategy switch (yes vs no) as predictors. Strategy switch was the only significant predictor (p = 0.007). People who switched study strategy recalled significantly more from L2 than those who did not change study strategy. The main effect of cue was not quite significant (p = 0.06) despite a large sample size. These findings largely replicate our previous research conducted at another institution (Sahakyan & Delaney, 2003), and they suggest that strategy switch provides a better explanation for the DF benefits than does the cue. However, it is important to note that we have always obtained cue effects both in prior research and in the current dataset, just not at the conventional levels of significance. This suggests that there is probably some escape from proactive interference that arises from context change. However, such effects are difficult to detect in overall recall rates (at least in a two-list design), and they are much smaller than the encoding effects arising from the strategy change.

7.3. Test Order and the Benefits of DF

One factor that could contribute to failures to detect DF benefits may be linked to test order. A recent meta-analysis on 20 studies that reported data on output order concluded that failure to detect DF benefits in many studies could be driven by the fact that L2 was tested after L1 in many prior studies (Pastötter et al., 2012). In addition to meta-analyses, Pastötter et al. (2012) conducted two studies, where recall order was manipulated, and they showed that the benefits of DF are much larger when L2 is tested first in the recall sequence than when it is tested after L1.

Although several studies from our lab controlled the encoding strategy and tested L2 after L1 (Sahakyan & Delaney, 2003; Sahakyan, Delaney, & Waldum, 2008), it is unlikely that our chosen test order could account for the absence of benefits in those studies. Sahakyan and Delaney (2010) counterbalanced test order while controlling study strategy for some participants and not for others. Although their published data were collapsed across test order, Figure 4.7 replots the data broken out by test order, and shows that the presence or absence of DF benefits was driven entirely by whether the study strategy was controlled or not; test order had no impact on the findings.

7.4. Summary

Most researchers now agree that DF benefits on L2 arise from a different mechanism than the costs. We argued that the underlying factor is the strategy change triggered by the forget cue. As with the context-change mechanism, the idea for this came from verbal reports given by participants

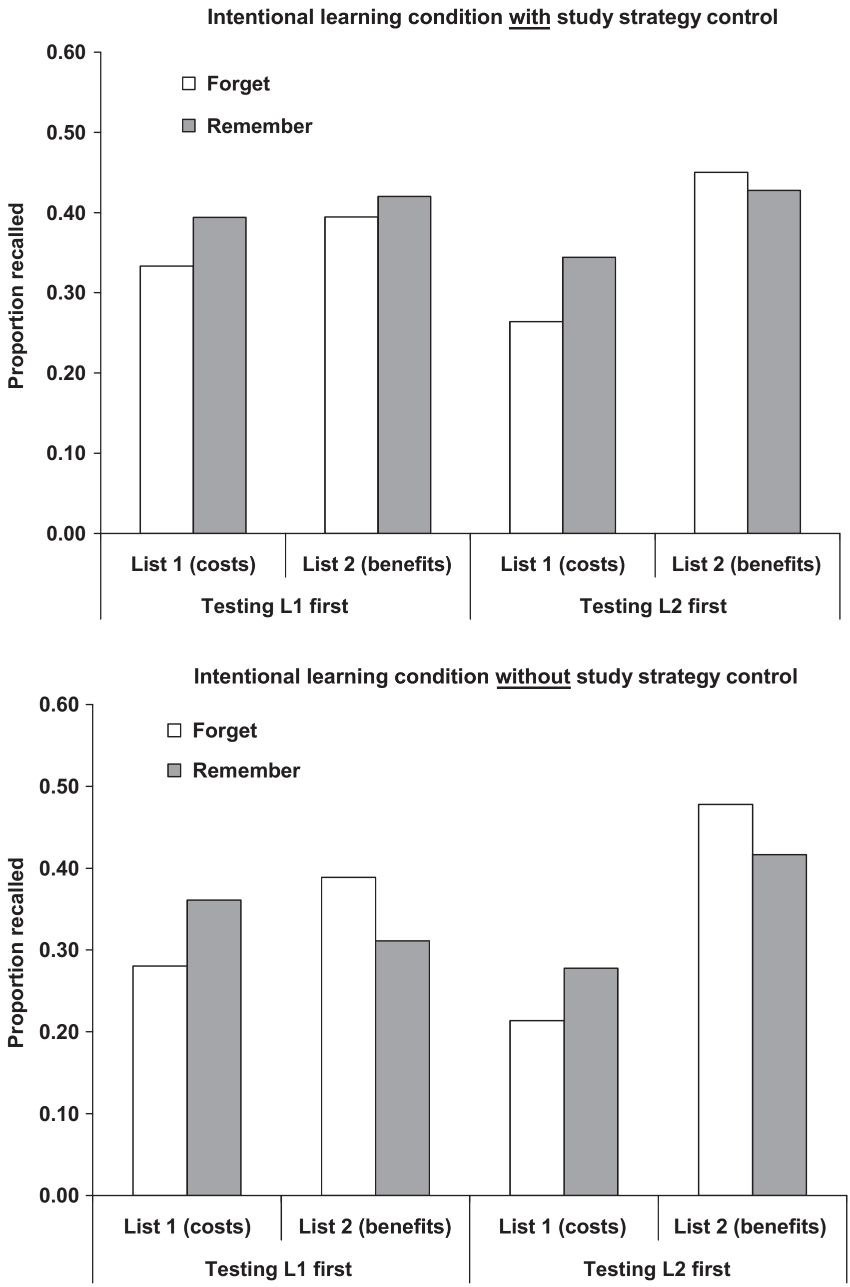

Figure 4.7 Directed forgetting costs and benefits as a function of test order, controlled study strategy (a), and uncontrolled study strategy (b). *(New figures based on data reported in Sahakyan and Delaney (2010))*.

in our experiments, and was supported by successful predictions in other experiments. For example, controlling participants' encoding strategy via instructions reduced or eliminated the benefits, as does the use of incidental learning procedures. This review also reports on a large-scale replication of the original work establishing the benefits were due to strategy changes: even though strategy switches are by no means common, they produce such large benefits to memory that they show up at the group level. We think that reducing build-up of proactive interference due to contextual isolation (the mechanism originally proposed by Sahakyan & Kelley, 2002 to explain the benefits) may also be important in some cases, but we have at best limited empirical evidence for that mechanism.

8. IMPLICATIONS FOR CLINICAL POPULATIONS

People with clinical disorders sometimes behave atypically on cognitive tasks. Demonstrations that clinical disorders alter behavior in cognitive tasks are useful because they can improve diagnosis (if the difference is sufficiently robust), differentiate subtypes of disorders, and clarify what functions are impaired in clinical populations. Cognitive difficulties can suggest treatments that may remediate symptoms, or predict the kinds of everyday problems that people may have. Of course, it is insufficient to know that a particular cognitive task is affected by a clinical diagnosis; we must have a theory as to the reason why the disorder affects behavior in the cognitive task. Hence, our theories about why cognitive effects occur are important in interpreting why clinical populations perform the task differently.

List-method DF is one cognitive task that people with clinical diagnoses sometimes perform differently from other people. Until the past decade, list-method DF deficits have generally been interpreted in terms of deficits in inhibition (for earlier reviews of the DF literature in clinical populations, see Cloitre, 1998; McNally, 2005). Significant DF is taken as evidence that inhibition is intact, while difficulties in DF are taken as evidence for impairments in inhibitory control. However, our review provides an alternative interpretation. Some clinical populations may have problems with context processing or context change. Although not everyone is convinced that the case for inhibition is well-established even in the other tasks (e.g. MacLeod et al., 2003; Raaijmakers & Jakab, 2013), many researchers assume that impairment on tasks like the stop-signal paradigm and RIF reflect inhibitory control, and that deficits on these tasks should be diagnostic of inhibitory deficits. If so, then if DF is also inhibitory in nature, we would expect that the same inhibitory deficits should occur there as well. We will review evidence that is consistent with a context-based interpretation of clinical difficulties in DF, and that other putatively "inhibitory" tasks do not show the same pattern of deficits as DF, consistent with our theory. The notion that DF impairments are linked to context processing (and not inhibition) provides new ways of thinking about how DF elucidates clinical disorders.

Having shown that context may be important to clinical disorders, we will next argue that many of the studies that have been conducted in the clinical literature are lacking in rigor, and that improving the methodological rigor of future studies will be necessary to draw firmer conclusions from DF research. We will provide specific advice for clinical researchers seeking to use DF as a tool for understanding context processing in clinical disorders. Without singling out any earlier study, we will suggest that many earlier studies could be repeated with more recently developed methodologies from the cognitive literature in order to yield more detailed conclusions about disorders.

8.1. Reinterpreting Clinical List-Method DF Results as Reflecting Context

Given that context is so important in memory theory, it is perhaps surprising that it is so underrepresented as an explanatory mechanism in clinical problems. Our ability to link events to their context plays an important role in understanding aging deficits, for example, where research suggests that binding items to their context may be extremely difficult for older adults. However, the same logic has rarely been applied to clinical disorders, where one might expect quite a lot of traction to be gained from linking disorders to mental context processing. In the next few sections, we will outline some disorders where we think that the context view may make sense of dissociations with other tasks that are assumed to reflect inhibition, and perhaps provide new ways of understanding how context is implicated in clinical problems.

8.1.1. Attention-Deficit Hyperactivity Disorder

An obvious place to begin the review is with attention-deficit hyperactivity disorder (ADHD), because it is widely believed that ADHD is in part a dysexecutive disorder (e.g. Barkley, 1997). A number of studies have argued that there are inhibitory deficits in ADHD patients, because they perform poorly in the think/no-think paradigm, stop-signal paradigm, and elsewhere (for a brief review, see Depue, Burgess, Wilcutt, Ruzic, & Banich, 2010).

Given that pattern of deficits, inhibitory theorists would likely expect that ADHD should impair list-method DF as well, but that does not appear to be true. Gaultney, Kipp, Weinstein, and McNeill (1999) compared children aged 8–15 who had ADHD and who did not. Using a standard list-method paradigm, they obtained both the costs and benefits of DF with both groups. Thus, ADHD children apparently show normal DF. White and Marks (2004) compared undergraduates with ADHD to non-ADHD undergraduates. Participants studied two lists of items using a procedure similar to Geiselman et al. (1983): Half of the items were designated as "learn" items that participants were supposed to memorize for a later test and the other half were "judge" items that they had to rate for pleasantness without intending to memorize them. The non-ADHD students replicated the results of DF costs reported by Geiselman et al. (1983)—namely, both learned and judged words suffered from DF. The results of the ADHD group, on the other hand, can be interpreted differently depending on how they are analyzed. The authors concluded that "individuals in the ADHD condition did not show reliable evidence of intentional forgetting." However, a reexamination of their data suggests a different conclusion: normal costs (11%) were present for the learn items among ADHD students, although for judge items the costs were absent (−6%). The pattern with judge words may be a false negative and may be worth following up in future work given that several studies have obtained DF in incidental encoding (e.g. Sahakyan & Delaney, 2005, 2010). The results with the learn words are largely consistent with the Gaultney et al. (1999) finding of intact DF in ADHD. In sum, while ADHD patients show deficits in other tasks that have been linked to inhibition, DF appears to be largely intact in this population.

8.1.2. Schizophrenia

Impairments in DF sometimes co-occur with deficits in other inhibitory tasks. For example, schizophrenia is thought to be associated with deficits in memory and inhibitory control (for reviews, see Hoff & Krennan, 2002; Perlstein, Carter, Barch, & Baird, 1998), and therefore researchers who believe that DF reflects inhibition should predict deficits in list-method DF in schizophrenics. Several studies have tested this possibility. In our judgment, these are among the best-conducted list-method DF studies in the clinical literature, and should be viewed as models for others.

Racsmány et al. (2008) examined medicated schizophrenics. While the control group showed robust costs and benefits, the schizophrenic patients showed neither costs nor benefits. Their overall memory was also lower.

Importantly, in a second experiment, they obtained robust RIF effects, suggesting dissociation between DF and RIF.

Soriano et al. (2009) examined medicated patients diagnosed with schizophrenia, schizoaffective disorder, and schizophreniform disorder. They were split according to whether they reported hallucinations on the Positive and Negative Syndrome Scale (PANSS). Although the omnibus statistics did not always work out due to low power, the planned comparisons suggested that the costs and benefits were present for both controls and nonhallucinators, whereas the patients who were hallucinating showed neither. Both groups of patients had impaired memory compared to controls. Thus, Soriano et al.'s (2009) findings suggest that the presence or absence of hallucinations is a critical variable in terms of whether schizophrenics show DF or not.

8.1.3. Generalized Anxiety Disorder

Few studies have found any executive functioning difficulties in generalized anxiety disorder (GAD). A recent study, for example, looked at postpartum women's GAD symptomology and found no correlation with a range of executive functions other than working memory capacity (Vadnais et al., 2012). Thus, there is not much evidence to suggest that GAD is associated with impaired inhibition. Nevertheless, a person with high anxiety may find it aversive to "think of something else" and may have difficulty encoding or shifting contexts. In other words, anxiety disorders may provide an example of difficulties shifting context despite intact inhibitory abilities. A study by Albu (2008) used the full design of the list method to examine GAD patients and patients without anxiety difficulties. Their L1 consisted of anxiety-related words and L2 consisted of neutral words. The major problem with the study is that the authors did not counterbalance which items were assigned to the forget and remember conditions. Therefore, any differences between the groups must be interpreted in light of possible differences due to item effects (we will discuss item effects more fully in Section 8.3). The non-GAD group showed robust costs (and, unusually, antibenefits). In contrast, the GAD group showed neither costs nor benefits. These findings suggest that the GAD patients might have a deficit in context change despite intact performance on other inhibitory tasks.

8.1.4. Depression

An even more interesting pattern would occur if a clinical problem improved DF. A study by Power, Dalgeish, Claudio, Tata, and Kentish (2000) examined subclinically dysphoric people, comparing people with low-Beck Depression

Inventory (BDI) scores to high-BDI scorers (dysphorics). In all their studies, they used mixed lists of positive and negative adjectives. Their Experiments 1 and 2 are difficult to interpret, as they did not include a remember control group or control for test order of the lists, but the "R–F" measure was positive and significant in both groups. However, their Experiment 3 used clinically depressed, high anxious, and healthy controls, and included a remember control group for each sample. The number of participants in each cell was extremely small (in some cases as few as five people), making interpretation difficult. Although most of their effects were in the correct direction for obtaining the costs and benefits, nothing was significant except an anti-forgetting effect for negative words among the depressed participants. This study suggests that negative words may be difficult to forget for depressed people, perhaps because they trigger rumination. Consistent with a bias to ruminate on negative words, Experiment 2 found that dysphorics recalled approximately equal numbers of positive and negative words, whereas the low-BDI participants showed a positivity bias and favored positive words. Indeed, negative words are difficult for depressives to update in working memory, which has led some authors to call them "sticky"—difficult to turn attention away from (Joormann, Levens, & Gotlib, 2011).

The Power et al. (2000) study therefore does not provide a clean test of whether depressives show greater DF than the controls. A recent poster by Lehman and Malmberg (2011b) provided a progress report on their large-scale study examining depressed peoples' DF, and suggested that depressives may be better able to change context. They used the full list-method DF design with two lists of neutral words, and found that clinically depressed people showed larger DF impairment than the control participants. Their argument was that depressed patients may have a built-in mechanism for producing more context change: depressive rumination. Although Joormann and Tran (2009) studied people high on rumination in a list-method study, they used a median split on a rumination scale, and did not find higher BDI scores among their ruminators. Hence, their failure to obtain enhanced DF among ruminators (using a no-control-group "R–F" statistic) may not be surprising.

8.2. Future Directions for Clinical Research Using the List Method

In sum, ADHD produces impaired executive functioning and poor performance on tasks thought to reflect inhibitory control, but nonetheless show intact DF. Contrastingly, GAD patients show impaired list-method

DF without clear executive or inhibitory deficits, suggesting a double dissociation between inhibitory/executive measures and DF. Schizophrenia produces impairments on both putatively inhibitory tasks and on DF (but see Racsmány et al., 2008 for dissociations between DF and RIF). Finally, depressed patients may show larger than normal DF, despite some evidence that they sometimes have impaired working memory. Taken together, the clinical studies are consistent with our argument that the basis of list-method DF is not inhibition but contextual change, and that disorders that affect the likelihood of effectively thinking of something else affect list-method DF.

These studies all focused on whether or not DF occurs in a particular population, but the obvious next step is to ask *why* DF is impaired. In this chapter, we have argued that DF is caused by context change, but also that context change is under volitional control. Impaired DF might reflect inability to deploy an effective strategy for forgetting. Perseverative difficulties could make it difficult to stop ongoing processing and deploy forgetting strategies, for example. It may be useful to include measures of what people do when attempting to forget (Foster & Sahakyan, 2011), as some people may fail to initiate context change spontaneously, but could do so normally with instruction (e.g. in the aging literature; Sahakyan, Delaney, & Goodmon, 2008). Alternatively, impaired DF could reflect difficulties in context processing, as was suggested for people with low working memory capacity (Section 5.12). Superior DF, likewise, could make sense if people are more able than usual to deploy an effective forgetting strategy—a case we will suggest might apply to depression (Lehman & Malmberg, 2011b). One possible method of getting at whether generalized context-change problems are typical of a disorder or whether it is a problem of initiating context change on one's own may be to use diversionary thought tasks like imagining your childhood home (Section 5.1).

That depressives seem more effective than others at forgetting neutral words (Lehman & Malmberg, 2011b) but not mixed lists of negative and positive words (Power et al., 2000) further suggests that forgetting may be dependent on the type of material being forgotten (McNally, Metzger, Lasko, Clancy, & Pitman, 1998). Different types of material could trigger thoughts that produce forgetting or could fail to do so. Studies contrasting forgetting of neutral and disorder-related words may be informative in other disorders as well, especially if they include appropriate remember control groups to facilitate differentiating between generalized memory biases and biases specific to forgetting.

8.3. Improving Research Design for Clinical Studies of List-Method DF

One problem with interpreting studies on list-method DF is that many of them are based on older research designs that provide data that are often misinterpreted. Hence, our next task is to suggest list-method designs that provide the best chance of accurately answering research questions about clinical populations' DF. Whether or not one accepts that DF is context-based, it is a good idea to use a design that can discriminate different theoretical possibilities. We have outlined our specific recommendations as Table 4.1.

8.3.1. Inclusion of a Separate Remember Condition

As noted in Section 2.1, a common design issue in DF studies is the absence of a remember control group. In such studies, the costs and benefits of DF have been assessed using the "R–F" measure, which can be a problematic assessment for DF effects. This measurement issue is particularly problematic when assessing DF in clinical populations. Specifically, if a clinical population is impaired on the costs but shows normal benefits, this would be impossible to distinguish using the "R–F" measure. According to the theory outlined in this chapter, smaller costs but normal benefits suggest difficulties with the use of context. Intact costs but impaired benefits suggest strategy-change issues. Thus, including a remember control group allows for a systematic assessment of the costs and benefits of DF.

It is understandable that including remember control groups when patient populations are scarce may be difficult. For this reason, many designs rely on comparing "R–F" in healthy vs patient populations. We think these studies are informative in that they indicate differences in DF between patients and healthy controls. However, one has to be very careful about interpreting them in terms of impaired inhibition (or even impaired context change), because both costs and benefits are rolled into the "R–F" differences. A possible solution is to adopt the four-list procedure (Section 2.2). Especially if

Table 4.1 Suggestions for Improved List-Method Directed Forgetting Designs

1. *Use a remember control group*—or at least the within-subjects design that includes both RR and FR lists
2. *Test list 1 and list 2 separately*—or at least calculate output percentiles
3. *Consider output order*—if a study mixes different types of items (e.g. trauma and neutral words), examine output order effects
4. *Avoid or acknowledge scaling effects*—baseline memory may differ across groups
5. *Avoid floor effects*—use appropriate numbers of items in each category

one is interested in the costs, there is little evidence that the within-subjects design distorts the effect, as reviewed earlier. It is important to counterbalance the assignment of words to lists as well; at least one study obtained odd results because they failed to ensure that the words assigned to the forget condition were assigned to the remember condition for other participants, and hence item effects emerged (Albu, 2008). The four-list procedure allows the researcher to compute both costs and benefits using a powerful within-subjects comparison, which is clearly advantageous. If there are concerns about order effects, they can also be addressed using this design. If there are serious order effects, one can also focus on only the first lists studied.

8.3.2. Controlling Test Order and Measuring Output Order

Another design issue is that most clinical studies have not controlled the test order of L1 and L2. In Section 2.1, we discussed the problem of output interference in the benefits. While research suggests that in healthy controls the costs are relatively unaffected by output interference effects (Pastötter et al. 2012), this may not necessarily be the case for patients. There could be systematic differences between patients and healthy controls as to whether they start recall from L1 or L2 for a variety of reasons.

Controlling test order may be particularly important when mixing several different types of items on the same list, as when one is interested in whether trauma or neutral words are harder to forget. Comparing the relative recall rate of different item types will be complicated by output interference, especially if different groups of participants output them in different orders. For example, if we mix trauma and neutral words on each list, then patients might begin recall from all the trauma words, whereas healthy controls might have no such bias. Such differences in output order would produce varying levels of output interference on different types of items across the groups, resulting in apparent differences in recall rates. Hence, we suggest that output order of the lists be controlled whenever possible.

At the very least, one should examine output order as a possible explanation for any observed results. One way to examine output order is to calculate the average output positions. The easiest way to do this is to create output percentiles, which examine the average position within the output of the items of a given type (e.g. Sahakyan, Delaney, & Waldum, 2008). For example, imagine we mix trauma and neutral words, and the output items are in the order *trauma*, *trauma*, *trauma*, *neutral*, *trauma*, *neutral*, *neutral*. The output percentile for trauma words would be $(1 + 2 + 3 + 5)/(7 \times 4) = 0.39$. For neutral words, it would be $(4 + 6 + 7)/(7 \times 3) = 0.81$. Thus, output for neutral words is occurring later than for trauma words for this participant. Further, significant differences in output percentiles between groups would be evidence for differences in output order of different types of items.

8.3.3. Scaling Effects

Another common problem when comparing groups of people is a scaling effect (also known as baseline effects). Some groups of people show general memory impairments, and hence have lower baseline memory. For example, this problem is well-known in the literature on aging—older adults have impaired recollection, and so they perform worse than younger adults on free recall tests. In this case, the costs might be numerically smaller in older adults than in younger adults due to reduced overall memory. If the costs are a percentage of the original memory, then Analysis of Variance (ANOVA) might incorrectly suggest that the costs are smaller for older adults than for younger adults. One solution is either to enhance the memory of the weaker group, such as by giving longer presentation times, or to provide more elaborate processing instructions before the forget cue, especially because neither study time nor depth of processing interacts with DF (Sahakyan, Delaney, & Goodmon, 2008; Sahakyan, Delaney, & Waldum, 2008). If the baseline recall rates are closer in the remember control group, then the differences in the costs may be more easily interpreted. Of course, if the costs are completely absent, it is more difficult to argue for a scaling effect (unless there is a floor or ceiling effect in the data).

8.3.4. Floor Effects

Many clinical studies also show floor effects in recall. The standard diagnosis for floor effects is that the confidence interval around the mean captures 0%. Another good indicator of floor effects is if a significant percentage of participants recall zero of a given type of item. When the range of responses is very small (e.g. 0, 1, 2 or 3), then a lot of participants would be needed to detect a meaningful difference in the means, as the measurement error is likely to exceed the size of the effect. Whenever there is an apparent floor effect, one should be prudent when interpreting mean differences, as the estimate of the variance is likely to be deflated, producing false positives.

It is difficult to avoid floor effects in recall, but a good start is to avoid very long lists; not more than 20 items are likely to yield reasonable recall rates. Likewise, if the items are fractionated into different types of items, such as traumatic and neutral words, then one should try to keep these sublists from getting too short. We suggest a minimum of eight items per category. Of course, this restricts the kinds of tests that can be run with a single design, but it is better to get meaningful answers about one or two different types of item than to get meaningless floor effects with a large number of different types of item.

8.4. Summary and Linking Clinical to Nonclinical Research on DF

While we have argued that the current theory can explain many of the differences in DF among clinical populations, in some ways, studies with special populations provide the best evidence for individual differences in DF. Even if contextual change can account for DF impairment in everyday people, it may be less successful among some clinical groups for a variety of reasons. The absence of DF in some clinical populations may provide clues as to why some seemingly healthy participants show little or no DF. Likewise, if some clinical populations show more effective DF than healthy participants, it may give us some insight into why some healthy participants show larger-than-usual DF effects. For example, it may well be that people who ruminate more or people who are fantasy-prone are better forgetters even in the absence of any clinical symptoms.

Clinical research could also shed light on the controversies surrounding what types of items are more or less forgettable. Some populations have difficulties related to specific types of words, which could enable us to better understand how differential processing interacts with the forget cue. Thus, more insights about the nature of DF can be gained through well-designed studies with clinical populations. Although basic research provides a foundation for interpreting population differences, research with special populations is equally necessary for the future development and refinement of DF theories. We are hopeful that by incorporating better, modern research designs that were developed in the cognitive literature to investigate DF that deeper conclusions can be drawn about context processing in various clinical populations. Repeating some of the earlier studies with these modern designs should be helpful in discriminating between different theoretical possibilities, and shed light on the nature of several important clinical disorders.

9. CONCLUDING THOUGHTS

Obtaining DF requires engaging in controlled strategies, and the decision to engage in those strategies may be mediated by beliefs about one's memory abilities or whether it is possible to intentionally forget. These findings highlight the importance of attending to participants' behavior and beliefs in understanding why memory phenomena occur. The costs and the benefits of DF can be observed or not depending on what participants choose to do. A closer attention to participant strategies is warranted in DF.

The context account as laid out here explains a substantial amount of DF data, while also serving as foundation for testable predictions regarding the impact of different manipulations on DF. The context-change mechanism is not uniquely suited to explain DF. Instead, the account brings DF into the family of effects that are explained by existing memory models, allowing us to make new testable predictions. These models and the rich empirical foundation of the environmental context literature serve as a foundation for new quantitative and qualitative predictions.

The traditional way of interpreting DF deficits in many special populations is that they have deficits in inhibitory control. The context account of DF, however, suggests new interpretations, opening up new directions for investigating context processing and/or contextual binding deficits in those populations.

Finally, a commonsense notion is that disruptions are inherently bad for memory. Work on DF and mental context-change paradigm suggests that how disruptive interruptions depend on the contents of the interruption and the extent to which they change mental context. Some interruptions are easier to recover from than others. A major challenge that lies ahead is to more fully understand why some tasks but not others create mental context change—in other words, to create a theory of what mental context consists of. We look forward to seeing how these theories develop in the next few decades. Understanding how mental context is represented and updated may turn out to be central to memory functioning in many more ways than we yet realize.

ACKNOWLEDGMENTS

The authors are grateful to Colleen Kelley, Colin MacLeod, and Tanya Jonker for their feedback and constructive comments on the earlier draft of this chapter.

REFERENCES

Albu, M. (2008). Automatic and intentional inhibition in patients with generalized anxiety disorder. *Cognition, Brain, & Behavior*, *12*, 233–249.

Anderson, M. C. (2003). Rethinking interference theory: executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445.

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: a consideration of three methods. In C. MacLeod & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 159–190). Tokyo: Springer-Verlag.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123.

Aslan, A., & Bäuml, K. T. (2008). Memorial consequences of imagination in children and adults. *Psychonomic Bulletin & Review*, *15*(4), 833–837.

Aslan, A., & Bäuml, K. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 264–269.

Aslan, A., Zellner, M., & Bäuml, K. T. (2010). Working memory capacity predicts listwise directed forgetting in adults and children. *Memory*, *18*(4), 442–450.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*, 65–94.

Barnier, A. J., Conway, M. A., Mayoh, L., Speyer, J., Avizmil, O., & Harris, C. B. (2007). Directed forgetting of recently recalled autobiographical memories. *Journal of Experimental Psychology: General*, *136*(2), 301–322.

Basden, B. H., & Basden, D. R. (1996). Directed forgetting: further comparisons of the item and list methods. *Memory*, *4*(6), 633–653.

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: a contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139–172). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

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*(3), 603–616.

Basden, B. H., Basden, D. R., & Morales, E. (2003). The role of retrieval practice in directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 389–397.

Basden, B. H., Basden, D. R., & Wright, M. J. (2003). Part-list reexposure and release of retrieval inhibition. *Consciousness and Cognition: An International Journal*, *12*(3), 354–375.

Bäuml, K. H. (2008). Inhibitory processes. In H. L. RoedigerIII (Ed.), *Cognitive psychology of memory* (pp. 195–220). Oxford: Elsevier.

Bäuml, K.-H., Hanslmayr, S., Pastotter, B., & Klimesch, W. (2008). Oscillatory correlates of intentional updating in episodic memory. *NeuroImage*, *41*, 596–604.

Bäuml, K. T., & Samenieh, A. (2010). The two faces of memory retrieval. *Psychological Science*, *21*(6), 793–795.

Bäuml, K. T., & Samenieh, A. (2012a). Influences of part-list cuing on different forms of episodic forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 366–375.

Bäuml, K. T., & Samenieh, A. (2012b). Selective memory retrieval can impair and improve retrieval of other memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 488–494.

Benjamin, A. S. (2006). The effects of list-method directed forgetting on recognition memory. *Psychonomic Bulletin & Review*, *13*, 831–836.

Benjamin, A. S., & Bjork, R. A. (2000). On the relationship between recognition speed and accuracy for words rehearsed via rote versus elaborative rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 638–648.

Bjork, R. A. (1970). Positive forgetting: the noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, *9*, 255–268.

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217–325). Washington, D. C.: V. H. Winston & Sons.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. Roediger & F. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition*, *5*, 176–196.

Bjork, E. L., & Bjork, R. A. (2003). Intentional forgetting can increase, not decrease, residual influences of to-be-forgotten information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 524–531.

Bjork, E., Bjork, R. A., & Anderson, M. C. (1998). Varieties of goal-directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103–137). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Flowerree symposium on cognition* (pp. 313–344). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Block, R. A. (1971). Effects of instructions to forget in short-term memory. *Journal of Experimental Psychology*, *89*(1), 1–9.

Cloitre, M. (1998). Intentional forgetting and clinical disorders. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 395–412). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Conway, M. A., & Fthenaki, A. (2003). Disruption of inhibitory control of memory following lesions to the frontal and temporal lobes. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *39*(4–5), 667–686.

Conway, M. A., Harries, K., Noyes, J., Racsmány, M., & Frankish, C. R. (2000). The disruption and dissolution of directed forgetting: inhibitory control of memory. *Journal of Memory and Language*, *43*(3), 409–430.

Craik, F. I., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 599–607.

Dalton, P. (1993). The role of stimulus familiarity in context-dependent recognition. *Memory & Cognition*, *21*(2), 223–234.

Delaney, P. F., & Knowles, M. E. (2005). Encoding strategy changes and spacing effects in the free recall of unmixed lists. *Journal of Memory and Language*, *52*(1), 120–130.

Delaney, P. F., & Sahakyan, L. (2007). Unexpected costs of high working memory capacity following directed forgetting and context change manipulations. *Memory & Cognition*, *35*, 1074–1082.

Delaney, P. F., Sahakyan, L., Kelley, C. M., & Zimmerman, C. A. (2010). Remembering to forget: the amnesic effect of daydreaming. *Psychological Science*, *21*(7), 1036–1042.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–478.

Depue, B. E. (2012). A neuroanatomical model of prefrontal inhibitory modulation of memory retrieval. *Neuroscience and Biobehavioral Reviews*, *36*(5), 1382–1399.

Depue, B. E., Burgess, G. C., Willcutt, E. G., Ruzic, L., & Banich, M. T. (2010). Inhibitory control of memory retrieval and motor processing associated with the right lateral prefrontal cortex: evidence from deficits in individuals with ADHD. *Neuropsychologia*, *48*, 3909–3917.

Dong, T. (1972). Probe versus free recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(5), 654–661.

Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 764–770.

Elmes, D. G., Adams, C., & Roediger, H. L. (1970). Cued forgetting in short-term memory: response selection. *Journal of Experimental Psychology*, *86*(1), 103–107.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*(3), 145–154.

Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition*, *13*(4), 333–345.

Foster, N. L., & Sahakyan, L. (2011). The role of forget-cue salience in list-method directed forgetting. *Memory*, *19*, 110–117.

Foster, N. L., & Sahakyan, L. (2012). Metacognition influences item-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1309–1324.

Gaultney, J. F., Kipp, K., Weinstein, J., & McNeill, J. (1999). Inhibition and mental effort in attention deficit hyperactivity disorder. *Journal of Developmental and Physical Disabilities*, *11*(2), 105–114.

Geiselman, R. E., & Bjork, R. A. (1980). Primary versus secondary rehearsal in imagined voices: differential effects on recognition. *Cognitive Psychology*, *12*(2), 188–205.

Geiselman, R. E., Bjork, R. A., & Fishman, D. (1983). Disrupted retrieval in directed forgetting: a link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, *112*, 58–72.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, *66*(3), 325–331.

Goernert, P. N., & Larson, M. E. (1994). The initiation and release of retrieval inhibition. *Journal of General Psychology*, *121*, 61–66.

Golding, J. M., & Gottlob, L. R. (2005). Recall order determines the magnitude of directed forgetting in the within-participants list method. *Memory & Cognition*, *33*(4), 588–594.

Golding, J. M., & Keenan, J. M. (1985). Directed forgetting and memory for directions to a destination. *American Journal of Psychology*, *98*, 579–590.

Golding, J. M., Long, D. L., & MacLeod, C. M. (1994). You can't always forget what you want: directed forgetting of related words. *Journal of Memory and Language*, *33*, 493–510.

Golding, J. M., & MacLeod, C. M. (1998). *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Gottlob, L. R., & Golding, J. M. (2007). Directed forgetting in the list method affects recognition memory for source. *Quarterly Journal of Experimental Psychology*, *60*(11), 1524–1539.

Hanczakowski, M., Pasek, T., & Zawadska, K. (2012). Context-dependent impairment of recollection in list-method directed forgetting. *Memory*.

Hanslmayr, S., Volberg, G., Wimber, M., Oehler, N., Staudigl, T., HYartmann, T., et al. (2012). Prefrontally driven downregulation of neural synchrony medicated goal-directed forgetting. *Journal of Neuroscience*.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review*, *11*(1), 125–130.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551.

Hoff, A. L., & Krennan, W. S. (2002). Is there a cognitive phenotype for schizophrenia: the nature and course of the disturbance in cognition? *Current Opinion in Psychiatry*, *15*, 43–48.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Hyde, T. S., & Jenkins, J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior*, *12*(5), 471–480.

Isarida, T., & Morii, Y. (1986). Contextual dependence of the spacing effect in free recall. *Japanese Journal of Psychology*, *57*(1), 20–26.

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 112–127.

Joormann, J., Levens, S. M., & Gotlib, I. H. (2011). Sticky thoughts: depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychological Science*, *22*(8), 979–983.

Joormann, J., & Tran, T. B. (2009). Rumination and intentional forgetting of emotional material. *Cognition and Emotion*, *23*(6), 1233–1246.
Joslyn, S. L., & Oakes, M. A. (2005). Directed forgetting of autobiographical events. *Memory & Cognition*, *33*(4), 577–587.
Kimball, D. R., & Bjork, R. A. (2002). Influences of intentional and unintentional forgetting on false memories. *Journal Of Experimental Psychology: General*, *131*, 116–130.
Koppel, R. H., & Storm, B. C. (2012). Unblocking memory through directed forgetting. *Journal of Cognitive Psychology*.
Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, *49*(1), 58–69.
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*(4), 970–988.
Lehman, M., & Malmberg, K. J. (2011a). Overcoming the effects of intentional forgetting. *Memory & Cognition*, *39*(2), 335–347.
Lehman, M., & Malmberg, K. J. (2011b, November). Modeling intentional forgetting in depressed participants. (Poster presentation at the 52nd annual meeting of the Psychonomic Society, Seattle, WA).
Liu, X. (2001). On the dynamics of directed forgetting: facilitation and interference in the updating of human memory. *Dissertation Abstracts International*, *61*(11), 6159B.
Macken, W. J. (2002). Environmental context and recognition: the role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 153–161.
MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 271–279.
MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah, NJ: Erlbaum.
MacLeod, C. M. (1999). The item and list methods of directed forgetting: test differences and the role of demand characteristics. *Psychonomic Bulletin & Review*, *6*(1), 123–129.
MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 43, pp. 163–214). New York, NY US: Elsevier Science.
Macrae, C., Bodenhausen, G.V., Milne, A. B., & Ford, R. L. (1997). On regulation of recollection: the intentional forgetting of stereotypical memories. *Journal of Personality and Social Psychology*, *72*(4), 709–719.
Malmberg, K. J., & Shiffrin, R. M. (2005). The 'One-Shot' hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 322–336.
Malpass, R. S., & Devine, P. G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology*, *66*(3), 343–350.
Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. (Vol. 1). New York: Academic Press.
Marsh, R. L., Meeks, J., Hicks, J. L., Cook, G. I., & Clark-Foos, A. (2006). Concreteness and item-to-list context associations in the free recall of items differing in context variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1424–1430.
McNally, R. J. (2005). Directed forgetting tasks in clinical research. In A. Wenzel & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research* (pp. 197–212). Washington, DC: American Psychological Association.
McNally, R. J., Metzger, L. J., Lasko, N. B., Clancy, S. A., & Pitman, R. K. (1998). Directed forgetting of trauma cues in adult survivors of childhood sexual abuse with and without posttraumatic stress disorder. *Journal of Abnormal Psychology*, *107*, 596–601.
Mensink, G., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434–455.

Minnema, M.T., & Knowlton, B. J. (2008). Directed forgetting of emotional words. *Emotion*, *8*(5), 643–652.

Mulji, R., & Bodner, G. E. (2010). Wiping out memories: new support for a mental context change account of directed forgetting. *Memory*, *18*, 763–773.

Nelson, D. L., Goodmon, L. B., & Ceo, D. (2007). How does delayed testing reduce effects of implicit memory: context infusion or cuing with context? *Memory & Cognition*, *35*(5), 1014–1023.

Nelson, D., & McEvoy, C. L. (2005). Implicitly activated memories: the missing links of remembering. In C. Izawa & N. Ohta (Eds.), *Human learning and memory: Advances in theory and application: The 4th Tsukuba International Conference on memory* (pp. 177–198). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*(2), 299–324.

Parker, A., Gellatly, A., & Waterman, M. (1999). The effect of environmental context manipulation on memory: dissociation between perceptual and conceptual implicit tests. *European Journal of Cognitive Psychology*, *11*(4), 555–570.

Pastötter, B., & Bäuml, K.-H. (2007). The crucial role of post-cue encoding in directed forgetting and context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 977–982.

Pastötter, B., & Bäuml, K. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 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. T. (2012). List-method directed forgetting: the forget cue improves both encoding and retrieval of postcue information. *Memory & Cognition*, *40*(6), 861–873.

Perlstein, W. M., Carter, C. S., Barch, D. M., & Baird, J. (1998). The Stroop task and attention deficits in schizophrenia: a critical evaluation of card and single-trial Stroop methodologies. *Neuropsychology*, *12*, 414–425.

Postman, L. (1964). Acquisition and retention of consistent associative responses. *Journal of Experimental Psychology*, *67*(2), 183–190.

Power, M. J., Dalgleish, T. T., Claudio, V. V., Tata, P. P., & Kentish, J. J. (2000). The directed forgetting task: Application to emotionally valent material. *Journal Of Affective Disorders*, *57*, 147–157.

Raaijmakers, J. G. W., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory and Language*, *68*, 98–122.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Racsmány, M., & Conway, M. A. (2006). Episodic inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 44–57.

Racsmány, M., Conway, M. A., Garab, E. A., Cimmer, C., Janka, Z., Kurimay, T., et al. (2008). Disrupted memory inhibition in schizophrenia. *Schizophrenia Research*, *101*(1–3), 218–224.

Racsmány, M., Conway, M. A., Garab, E. A., & Nagymáté, G. (2008). Memory awareness following episodic inhibition. *Quarterly Journal of Experimental Psychology*, *61*(4), 525–534.

Reitman, W., Malin, J. T., Bjork, R. A., & Higman, B. (1973). Strategy control and directed forgetting. *Journal Of Verbal Learning & Verbal Behavior*, *12*, 140–149.

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*(2), 261–269.

Roediger, H., & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Roediger, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 91–105.

Russo, R., Ward, G., Geurts, H., & Scheres, A. (1999). When unfamiliarity matters: changing environmental context between study and test affects recognition memory for unfamiliar stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 488–499.

Sahakyan, L. (2004). Destructive effects of 'forget' instructions. *Psychonomic Bulletin & Review*, *11*(3), 555–559.

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*, 195–201.

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., & Delaney, P. F. (2010). Item-specific encoding produces an additional benefit of directed forgetting: evidence from intrusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1346–1354.

Sahakyan, L., Delaney, P. F., & Goodmon, L. B. (2008). "Oh, honey, I already forgot that": strategic control of directed forgetting in older and younger adults. *Psychology and Aging*, *23*, 621–633.

Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor in strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review*, *11*, 131–134.

Sahakyan, L., Delaney, P. F., & Waldum, E. R. (2008). Intentional forgetting is easier after two 'shots' than one. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 408–414.

Sahakyan, L., & Foster, N. L. (2009). Intentional forgetting of action: comparison of list-method and item-method directed forgetting. *Journal of Memory and Language*, *61*, 134–152.

Sahakyan, L., & Goodmon, L. B. (2007). The influence of directional associations on directed forgetting and interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 1035–1049.

Sahakyan, L., & Goodmon, L. B. (2010). Theoretical implications of extralist probes for directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 920–937.

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., Waldum, E. R., Benjamin, A. S., & Bickett, S. P. (2009). Where is the forgetting with list-method directed forgetting in recognition? *Memory & Cognition*, *37*(4), 464–476.

Shapiro, S., Lindsey, C., & Krishnan, H. (2006). Intentional forgetting as a facilitator for recalling new product attributes. *Journal of Experimental Psychology: Applied*, *12*(4), 251–263.

Sheard, E. D., & MacLeod, C. M. (2005). List method directed forgetting: return of the selective rehearsal account. In N. Ohta, C. M. MacLeod & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 219–248). Tokyo: Springer.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 460–471.

Smith, E. E., Barresi, J., & Gross, A. E. (1971). Imaginal versus verbal coding and the primary-secondary memory distinction. *Journal of Verbal Learning and Verbal Behavior*, *10*(6), 597–603.

Smith, S. M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology*, *6*, 125–139.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203–220.

Soriano, M. F., & Bajo, M.T. (2007). Working memory resources and interference in directed-forgetting. *Psicológica*, *28*, 63–85.

Soriano, M. F., Jiménez, J. F., Román, P., & Bajo, M. T. (2009). Intentional inhibition in memory and hallucinations: directed forgetting and updating. *Neuropsychology*, *23*, 61–70.

Spillers, G. J., & Unsworth, N. (2011). Are the costs of directed forgetting due to failures of sampling or recovery? Exploring the dynamics of recall in list-method directed forgetting. *Memory & Cognition*, *39*(3), 403–411.

Spitzer, B., & Bäuml, K. (2007). Retrieval-induced forgetting in item recognition: evidence for a reduction in general memory strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 863–875.

Spitzer, B., & Bäuml, K. (2009). Retrieval-induced forgetting in a category recognition task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 286–291.

Steyvers, M., & Malmberg, K. J. (2003). The effects of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 760–766.

Storm, B. C., & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition*, *40*(6), 827–843.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford.

Tulving, E., & Arbuckle, T.Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72*, 145–150.

Unsworth, N., & Engle, R. W. (2008). Speed and accuracy of accessing information in working memory: an individual differences investigation of focus switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 616–630.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Dynamics of context-dependent recall: an examination of internal and external context change. *Journal of Memory and Language*, *66*(1), 1–16.

Vadnais, S. A., Behm, A., Laake, L. M., Lopez, N. M., Oddi, K. B., Wu, K., et al. (2012, May). Executive function correlates of symptoms of specific anxiety disorders and major depression. (Poster presented at the 24th Annual Convention of the Association for Psychological Science). Chicago, IL.

Veling, H., & van Knippenberg, A. (2004). Remembering can cause inhibition: retrieval-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 315–318.

Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition*, *32*, 1265–1272.

Waldum, E. R., & Sahakyan, L. (2012). Putting congeniality effects into context: investigating the role of context in attitude memory using multiple paradigms. *Journal of Memory and Language*, *66*(4), 717–730.

Wessel, I., & Merckelbach, H. (2006). Forgetting "murder" is not harder than forgetting "circle": listwise-directed forgetting of emotional words. *Cognition and Emotion*, *20*, 129–137.

Whetstone, T., Cross, M. D., & Whetstone, L. M. (1996). Inhibition, contextual segregation, and subject strategies in list method directed forgetting. *Consciousness and Cognition*, *5*, 395–417.

White, H. A., & Marks, W. (2004). Updating memory in list-method directed forgetting: individual differences related to adult attention-deficit/hyperactivity disorder. *Personality and Individual Differences*, *37*, 1453–1462.

Wilson, S., Kipp, K., & Chapman, K. (2003). Limits of the retrieval-inhibition construct: list segregation in directed forgetting. *Journal of General Psychology*, *130*(4), 341–358.

Zellner, M., & Bäuml, K. (2006). Inhibitory deficits in older adults: list-method directed forgetting revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 290–300.