The long and the short of it? An explanation of the spacing effect over long time scales.

A thesis

submitted in partial fulfilment

of the requirements for the Degree of

Doctor of Philosophy

at

Otago University

by

Christopher Dimick Smith

Otago University

2018

Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

The long and the short of it? An explanation of the spacing effect across long

by

Christopher Dimick Smith

The spacing effect is the observation that repetitions spaced out in time produce better learning and retention than repetitions massed closer together in time. There has been a long history of research on the spacing effect but there is not currently a satisfactory explanation for why it occurs. Furthermore, while empirical data on the spacing effect across long timescales has been accumulating, this data has not been integrated into the theoretical explanations of the effect. In this thesis I explore a new theory of why the spacing effect occurs, one based on memory reconsolidation. In experiments 1 to 3, I test paradigms for their suitability for testing a prediction of the reconsolidation account of the spacing effect. Unfortunately, the findings from previous studies did not replicate, making it impossible to probe the reconsolidation account of the spacing effect using these paradigms. In Experiment 4a I found a suitable paradigm and in Experiment 4b I set about testing a prediction of the reconsolidation account. Specifically, that manipulating the strength of a memory should influence the reconsolidation process and, as a consequence, the spacing effect. The results of experiment 4b were consistent with this prediction. In Experiments 5a and 5b I tested a second prediction of the reconsolidation account of the spacing effect. Specifically, that different mechanisms underly the spacing effect over short and long timescales. Consistent with the reconsolidation account of the spacing effect it appears retrieval difficulty influences the spacing effect over short timescales but not long timescales. I conclude by exploring future experiments that should be conducted in order to further test the reconsolidation account and the functional significance of the spacing effect.

Keywords: Spacing effect, reconsolidation, memory consolidation, distributed practice, learning, retention, forgetting.

Table of Contents

Acknowledgements

I would like to thank my supervisor Damian Scarf who has been a great source of support and good humour throughout the four years of my doctorate work. I would also like to thank my parents, Nicola and Julian and my grandparents Donald Smith, Joan and Trevor Dimick for their love and support. I would also like to thank the members of the ABCD lab for contributing to a great environment throughout my PhD years.

1 Chapter 1: Observations on the spacing effect

1.1 Introduction

Throughout history there has been much speculation on the nature of memory. In the 1880's Herman Ebbinghaus, a German philosopher, sought to move beyond mere speculation and conducted one of the earliest investigations in experimental psychology (Ebbinghaus, 1885/1964). Ebbinghaus employed himself as the sole participant and set about learning lists of nonsense syllables by repeating them out loud to himself. After a certain number of repetitions he would test himself by attempting to recall a list without errors. From this investigation Ebbinghaus made a number of observations that have stood the test of time. One of his observations was that, when he learned a 12-item list of nonsense syllables, if he spaced his repetitions across 3 days, it took him 38 repetitions to reach his standard for learning. However, if he massed his repetitions all in a single day it took him 68 repetitions to reach the same standard (Ebbinghaus, 1885/1964). This observation that it was more effective for repetitions to be spaced out across time became known as the *spacing effect* or the *distributed practice effect*. The effect has been a topic of much experimental investigation since Ebbinghaus first reported it and efforts to understand the mechanism(s) that underlie the effect continue to this day. As one psychologist eloquently put it "fads come and go in psychology, but research on the spacing effect has withstood the test of time and significance" (Dempster, 1989, pg 310).

After Ebbinghaus published his findings, one of the first questions raised was whether spaced repetitions were more effective because the learner was less fatigued during the learning process relative to when repetitions were massed in a single session (McGeoch & Irion, 1952). To test this, Jost (1897, as cited in Youtz, 1941) intermixed spaced and massed repetitions so that the level of fatigue was comparable across conditions. Despite fatigue being equalised he found that spaced repetitions resulted in better learning than massed repetitions. Other evidence from that time also

suggested that fatigue was unlikely to be a significant factor. For example, some studies employed very short periods of learning for the massed group that were very unlikely to cause fatigue (Easley, 1937; Hovland, 1938). Finally, Renshaw and Schwarzbek (1938a) examined the effect of different degrees of rest or spacing periods at different points in the learning process. If the spacing effect was due to reduced fatigue, you would expect a rest would be more effective later in the learning process given, in theory, there would be a greater build-up of fatigue. They found instead, that rests were more effective early in the learning process (Renshaw & Schwarzbek, 1938a, 1938b).

1.2 The spacing effect over short timescales[1]

Surprisingly, despite the fact Ebbinghaus employed long time scales in his research, the vast majority of research that followed his work has been conducted using very short spacing intervals in the order of seconds and minutes (Cepeda et al., 2006). The standard method of investigation involves presenting participants with a list of words or word-pairs. Within the list, individual words vary in the frequency with which they appear, with some words being presented once and others two or more times (e.g., Glenberg, 1976; Glenberg & Lehmann, 1980; Melton, 1967; Shaughnessy, Zimmerman, & Underwood, 1972; Young, 1971). For the words presented two or more times, the space between the presentation is manipulated. Some are massed, with the repetition occurring immediately after its first presentation, and others are spaced, with a varying number of words (or lag) between each presentation.

[1] The following observations on the spacing effect are divided into short and long timescales for several reasons. Firstly, some reviews of the spacing effect have not included studies across long timescales (Dempster, 1989; Greene, 2014; Hintzman, 1974) and other reviews that have included both short and long timescales have ignored interesting observations about the spacing effect over long timescales due to the number short studies greatly outweighing the number of long studies included (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Donovan & Radosevich, 1999; Janiszewski, Noel, & Sawyer, 2003; Moss, 1996). Secondly, as will be discussed in chapter 2 and 4 there may be different mechanisms that explain the spacing effect over short and long timescales.

For example, after the first presentation the word "mammal" there might be 6 different words presented between this first presentation and the repetition of the word in the list. Following a delay, memory for all words is tested. The spacing effect observed using this methodology is often called the *lag effect* and it is easiest to denote this methodology as the lag paradigm.

One significant question explored using the lag paradigm was whether the spacing effect is due to spaced items being rehearsed more than massed items. While differential rehearsal can, under certain conditions, produce what Delaney, Verkoeijen, and Spirgel (2010) term an "impostor spacing effect", the spacing effect is still observed when rehearsal is controlled so that equal rehearsal is given to massed and spaced items (Delaney & Verkoeijen, 2009; Hovland, 1938, 1940; Kahana & Howard, 2005; Shaughnessy et al., 1972; Zimmerman, 1975). Additionally, a spacing effect is observed in children too young to rehearse words and for pictorial materials that are extremely difficult for adults to rehearse (Hintzman & Rogers, 1973; McGeoch & Irion, 1952; Rea & Modigliani, 1987; Toppino, 1991; Toppino, Kasserman, & Mracek, 1991). The evidence ruling out fatigue and differential rehearsal explanations of the spacing effect has led researchers to develop explanations based on the memory processes engaged during learning and repetitions. Some of these explanations are discussed in Chapter 2.

There are a number of experimental findings that have come from research employing the lag paradigm that would need to be accounted for in any mechanistic explanation of the spacing effect. One important finding is the inverted-U relationship between spacing and retention (Glenberg, 1976, 1977; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963; Verkoeijen, Rikers, & Schmidt, 2005). That is, initially as the spacing of repetitions increases recall improves but, after a certain inflection point, further increases in the spacing between repetitions results in *worse* recall and the downward portion of the inverted-U (Figure 1). Additionally, the shape of the curve changes depending on the delay between the final repetition and the retention test. In particular, as the retention delay increases, the

optimal spacing interval changes and also becomes longer (Figure 1). Moreover, when there is a very short retention interval (e.g., an immediate test) massing is generally superior to any kind of spacing.

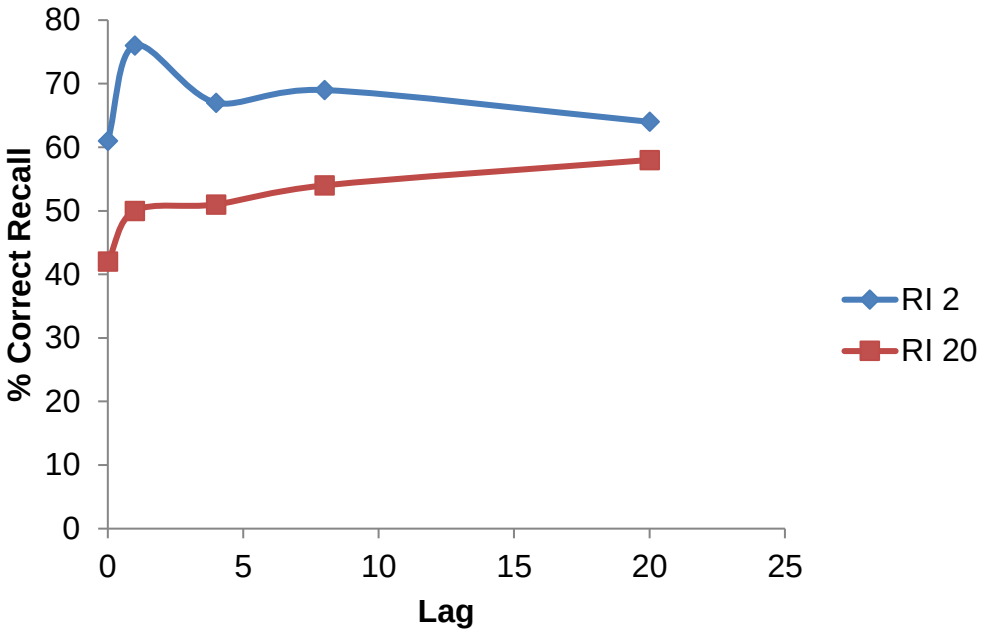

Figure 1. Mean percentage correct recall for word-pairs as a function of lag and retention interval. Data points redrawn from Balota, Duchek, and Paullin (1989). RI = Retention interval. Lag = the number of items between presentations.

Another observation that must be accounted for when explaining the spacing effect is study-phase retrieval (Asch, 1969; Delaney et al., 2010; Wahlheim, Maddox, & Jacoby, 2014). Study-phase retrieval is the technical term for participants recognising that they have studied the repeated item previously. A necessary component of this recognition process is that participant's retrieve from memory information about the first presentation, such as temporal and content information (Hintzman, Summers, & Block, 1975; Thios & D'Agostino, 1976). There are several pieces of evidence that support the importance of study-phase retrieval for the spacing effect. One piece of evidence is that, when the experimenter prevents study-phase retrieval, the spacing effect no longer occurs despite the fact the repeated item is still processed (Asch, 1969; Thios & D'Agostino, 1976). Wahlheim et al. (2014) provided another piece to this puzzle by manipulating the degree to which participants recognised paired-associates as a repetition. Wahlheim et al. (2014) found that increasing participant's

recognition that the presentation of a paired-associate was a repetition (i.e., enhancing retrieval of the first occurrence of the paired-associate) enhanced the spacing effect. This seems to suggest that part of the spacing effect is a combination or binding together of two experiences, in this case the first and second repetition of the paired-associate.

Related to this idea is the observation that spaced repetitions lead to "superadditive" recall (Begg & Green, 1988; Benjamin & Tullis, 2010). In essence, additive recall is the probability of recall we would expect if two presentations of an item were cognitively independent. It is calculated with the equation: $p_1 + (1-p_1)*p_2$. In this case p_1 is the probability of recall based on the first presentation of a word-pair and p_2 is the probability of recall based on the second presentation of a word-pair. To understand additivity one could simply imagine a probability tree (Figure 2). We might suppose that a single presentation provides a 0.5 probability of correct recall at test. The first scenario of correct recall is the trace formed from p_1 when it is successfully retrieved and the probability of this occurring is 0.5. The alternative scenario for correct recall is that there is a failure to retrieve the trace formed from p_1 (1 - 0.5) but there is retrieval of the trace formed by presentation 2 (with a probability of 0.5). This leads to (1 – 0.5)*0.5 = 0.25. The additive recall is therefore calculated by adding together the two scenarios for correct recall 0.5 + 0.25 = 0.75. Superadditive recall is when the recall based on two or more presentations is *greater* than the calculated additive recall. Using the example provided, recall is superadditive if the observed probability of recall is greater than 0.75. Superadditivity has been observed with some degree of generality in the spacing effect literature, with studies showing it occurs in free-recall, recognition, and cued-recall (Benjamin & Tullis, 2010). Superadditivity suggests that the two experiences are combined in some way and, in a sense, means that the whole is greater than the sum of the parts.

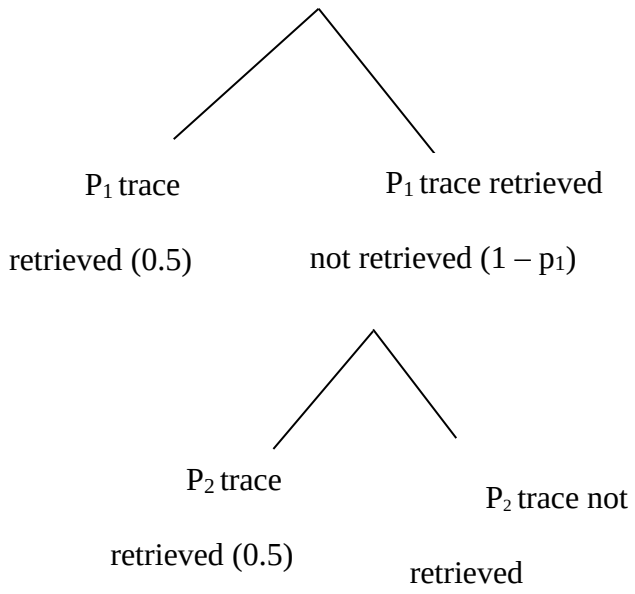

Figure 2. Probability tree for correct recall of an item if the traces formed by two presentations are cognitively independent.

The lag paradigm generally uses word lists, paired-associates, or pictures as stimuli for learning (e.g., Hintzman & Rogers, 1973; Paivio, 1974; Toppino, 1993). The spacing effect, however, has also been observed for learning and retaining skills (Donovan & Radosevich, 1999; Lee & Genovese, 1988). In these tasks, instead of using intervening items to create the spacing interval, participants are generally given time to rest. There are several notable observations of the spacing effect for skill tasks over short timescales. The first observation is that spacing enhances learning and creates temporary performance effects (Abrams & Grice, 1976; Donovan & Radosevich, 1999; Metalis, 1985). A particularly illustrative example is a study by Bourne and Archer (1956). They had participants learn a rotary pursuit task. In this task participants are given a stylus and a rotating wheel with a target on it. The participant's job is to keep the stylus on the target. Bourne and Archer (1956) compared participants who had 0, 15, 30, 45 and 60 seconds of rest between trials. As shown in Figure 3, throughout acquisition the amount of spacing influenced participants' performance. The best performance occurred in the group with 60 seconds of rest and the worst performance occurred in the

group with 0 seconds of rest. The remaining groups displayed intermediate performance during acquisition. The fact that the superiority of spacing is maintained after a 5-minute break indicates that spacing influenced learning and not just participants' performance. However, the fact that the groups are closer together after a 5-minute rest indicates that spacing also has a temporary influence on performance during acquisition. The conclusion that spacing has temporary performance effects is further supported in several other studies that find that spacing influences acquisition, but after a break there is no difference in the performance of the spaced and massed groups (Garcia, Moreno, Reina, Menayo, & Fuentes, 2008; Stelmach, 1969).

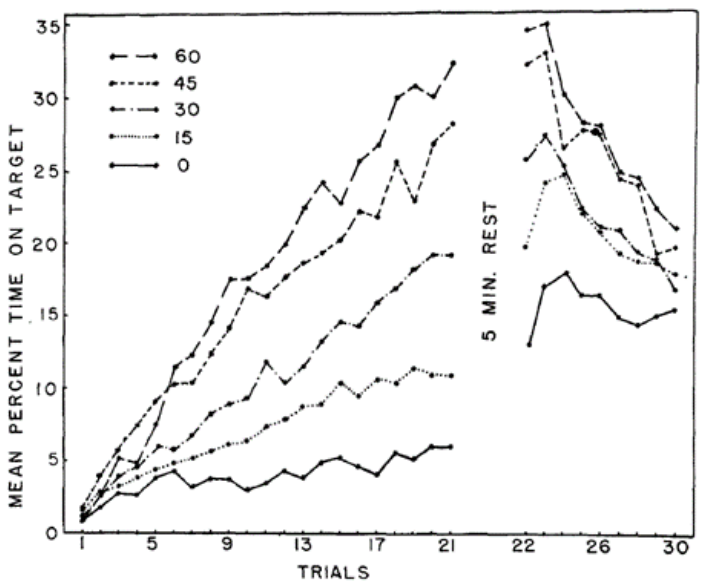

Figure 3. Mean percentage of time on target for different amounts of trial spacing in rotary pursuit. Figure taken from Bourne and Archer (1956).

A second observation about the spacing effect in skill tasks over short timescales is that the effect is influenced by the overall complexity of the task. More specifically, Donovan and Radosevich (1999) found a large effect size (d = 0.97) for spacing in tasks that were low in overall complexity (e.g., rotary pursuit, ball toss, etc.) and a trivial effect size (d = 0.07) for tasks that were high in overall complexity (e.g., air traffic controller simulation, musical performance, etc.). There has been little empirical exploration of why this observation occurs, but it is possibly due to the fact that extended periods of concentration are important in more complex tasks and the spacing disrupting this process.

1.3 The spacing effect over long time-scales

1.3.1 The spacing effect in skill-related tasks

1.3.1.1 Adults

Studies utilizing skill learning tasks to investigate the spacing effect using long spacing intervals have, by and large, compared a group that completes all of its training within one day (i.e., massed) to a group that completes its training across multiple days (i.e., spaced). As one might expect, spaced practice generally leads to better learning[2] and retention than massed practice (Arthur Jr et al., 2010; Dail & Christina, 2004; Shea, Lai, Black, & Park, 2000). Additionally, spacing has proven beneficial for a very wide range of skills such as playing video games (Shebilske, Goettl, Corrington, & Day, 1999; Stafford & Dewar, 2014), interviewing (Heidt, Arbuthnott, & Price, 2016), learning surgical skills (Andersen, Mikkelsen, Konge, Cayé-Thomasen, & Sørensen, 2016; Spruit, Band, & Hamming, 2014; Verdaasdonk, Stassen, Van Wijk, & Dankelman, 2007), playing a piano sequence (Rubin-Rabson, 1940; Simmons, 2011), balancing on a swaying platform (Shea et al., 2000), electrical testing of a vehicle charger (Hagman, 1980), learning to enhance alpha waves through bio-feedback (Albert, Simmons, & Walker, 1974) and golf putting (Dail & Christina, 2004).

1.3.1.1.1 Intensity of training

Most of the studies that investigate the spacing effect for skill-related tasks compare a group that completes their sessions across multiple days to one that completes all sessions within a single day. A smaller number of studies have investigated whether, when training is spaced over multiple days,

[2] Learning in this case is defined as performance at the end of training or soon after training has been completed and retention is defined as performance after a delay.

manipulating the intensity of training influences learning and retention. For example, is 1 hour of training per day for 16 days better than the more intense schedule of 4 hours of training per day for 4 days? Baddeley and Longman (1978) addressed this question by varying the number of sessions per day (1 or 2) and the number of hours in a session (1 or 2) that postmen were trained to type. After approximately 60 hours of training, postmen who completed the least intense training (1 session of 1 hour per day) learned to type faster and more accurately than the other more intense groups (Figure 4A). Additionally, a series of retention tests conducted 1, 3 and 9 months after training revealed a somewhat less clear-cut result, this was partly due to the least intense group going on holiday and completing less training than the other groups, but the overall trend was for the group with the most intense training (2 sessions of 2 hours per day) to perform worse than the other groups (Figure 4B). There are some other studies, though generally less systematic, that are consistent with Baddeley and Longman's (1978) finding that less intense training, spread across a larger number days, provides better learning (De Win, Van Bruwaene, De Ridder, & Miserez, 2013; Kauffeld & Lehmann-Willenbrock, 2010; Knapp & Dixon, 1950; Knapp, Dixon, & Lazier, 1958; Lashley, 1915; Ruch, 1928).

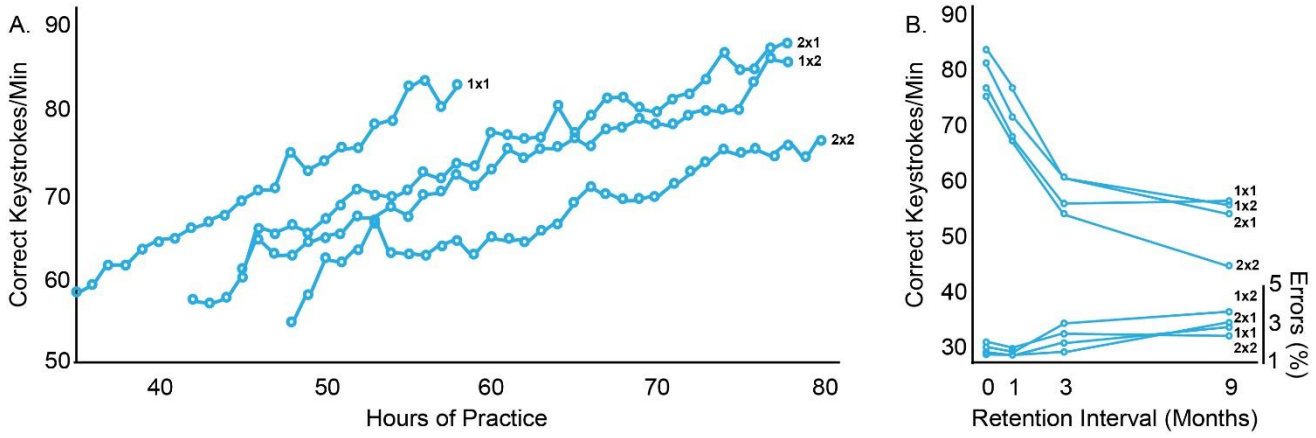

Figure 4. Learning and retention performance of postmen learning to type from Baddeley and Longman (1978). 1x1: one session of 1 hour per day. 1x2: one session of 2 hours per day. 2x1: two sessions of 1 hour per day. 2x2: two sessions of 2 hours per day.

Obviously, despite Baddeley and Longman's (1978) finding that the least intense training resulted in the greatest learning, there are likely lower limits on training intensity after which

performance declines. For example, Paik and Ritter (2015) had participants learn to balance an inverted pendulum under one of four different practise schedules. For our purposes the four practice schedules can be characterised from least intense to most intense. Paik and Ritter (2015) found that the intermediate intensity group (the hybrid-massed group), who completed 8 sessions across 1 week, showed significantly better learning than the least intense group (the spaced group) who completed 8 sessions across 2 weeks.

Another approach to investigating the intensity of daily sessions is by varying the total number of trials per day. Studies investigating perceptual and visuo-motor learning have found that a minimum number of trials is required for learning to occur (Aberg, Tartaglia, & Herzog, 2009; Wright & Sabin, 2007), and that there is an optimal number of trials per day for learning and going beyond this optimum produces minimal additional learning (Goedert & Miller, 2008; Molloy, Moore, Sohoglu, & Amitay, 2012; Savion-Lemieux & Penhune, 2005). For example, Wright and Sabin (2007) had participants learn to either discriminate the frequency of tones or the time interval

Table 1.
Summary of the main studies discussed that manipulate the intensity of training.

Study	Spacing groups	Task	Result
Baddeley and Longman (1978)	1 session of 1 hour per week day, 2 sessions of 1 hour per week day, 1 session of 2 hours per week day, 2 sessions of 2 hours per week day.	Postal office typing system	1 session of 1 hour per day showed better learning and retention.
Paik and Ritter (2015)	Spaced: 8 sessions across two weeks. Massed: 8 sessions across two days. Hybrid Massed: 8 sessions across 1 week.	Balancing an inverted pendulum	Hybrid Massed showed better learning and retention test performance than the other groups.
Wright and Sabin (2007)	360 trials per day or 900 trials per day	Tone discrimination: interval between tones, tone frequency.	Tone interval: 360 trials per day resulted in more efficient learning. Tone frequency: 900 trials per day resulted in better learning.
Young (1954)	4 days per week vs 2 days per week (same total number of sessions for each schedule).	Archery or badminton	Task dependent: archery improved more completing 4 days per week, badminton improved more completing 2 days per week.
Harmon and Miller (1950)	1 session per week, 3 sessions per week, expanding spacing interval (1,2,3,5,8…), 1 session per day…	Billiards	The group with the expanding spacing interval showed the showed the highest level of learning.

between tones, for 360 or 900 trials per day. For discriminating the frequency of tones, participants who received 360 trials per day failed to improve above their baseline level of performance while participants who received 900 trials per day displayed consistent improvement. In contrast, when discriminating the interval between tones, participants in both groups displayed the same rate of learning, indicating that going beyond 360 trials per day had no impact on this particular discrimination.

When learning a new sport or skill many people will practice for just a few days per week. It is, therefore, theoretically and practically interesting to understand the spacing effect when learning occurs on a weekly basis. Young (1954) had college students learn and practice badminton or archery 2 or 4 days per week. For badminton, the students improved more when they practiced 2 days per week compared to 4 days per week. In contrast, the archery students improved more when they practiced 4 days per week compared to 2 days per week. Young (1954) speculated that the results were due to differences in participants' prior experience of skills related to the two sports. Many participants probably had prior experience in racket sports and this meant they could improve their badminton skills with relatively spaced sessions, but for archery they required more concentrated sessions to build up their basic skills.

Harmon and Miller (1950) had college women who had no prior experience playing billiards learn and practice with different schedules, for a total of 9 sessions. They compared 4 different schedules: Group 1 completed 3 sessions per week for 3 weeks, Group 2 completed the 9 sessions across 9 consecutive days, Group 3 completed sessions across a gradually increasing interval (i.e., they practiced Day 1, 2, 3, 5, 8, 13, 21 and 34), and Group 4 completed 1 session per week for 9 weeks. At the end of training Group 3 performed significantly better than the other three groups. Somewhat similar to Young's (1954) explanation noted above, Harmon and Miller (1950) attributed the better performance of Group 3 to participants initially benefiting from concentrating their sessions to reach a

certain threshold of learning and then benefiting from the spacing of sessions to further improve performance.

Finally, there are some studies which do not find a spacing effect. These studies use a variety of tasks such as pattern tracing (Franklin & Brozek, 1947), prosthesis training (Romkema, Bongers, & van der Sluis, 2015) and assembling a machine gun (Schendel & Hagman, 1982). Some of these studies potentially reflect the fact that the tasks used are less sensitive to manipulations of training intensity because the training for all of the groups in these studies occurs over a long period of time and the tasks used are different to those reported on earlier (Franklin & Brozek, 1947; Massey, 1959; Murphree, 1971; Romkema et al., 2015). A few other studies potentially do not find a spacing effect because prior experience leads to a fast rate of learning and/or very little forgetting (Mitchell et al., 2011; Schendel & Hagman, 1982).

Overall, less intense daily training where learning is distributed over a larger number of days enhances learning and retention compared to more intense daily training. However, a certain minimum threshold of experience seems to be necessary for learning to occur in these daily sessions. This threshold varies depending on the type of task. Additionally, it would be useful to see whether the beneficial effect of gradually expanding the spacing interval found by Harmon and Miller (1950) could be replicated for billiards and other tasks to furnish theoretical accounts of the spacing effect and provide an effective schedule for learning and retaining skills. Finally, previous knowledge and experience is a potentially important variable for influencing precisely which spacing schedule is most beneficial for a particular skill.

1.3.1.1.2 Task complexity

As mentioned earlier, in their review of the spacing effect, Donovan and Radosevich (1999) found that for tasks categorized as highly complex (e.g., aeroplane control simulation) the effect size of the spacing effect was very small (d = 0.07). The implication of this finding is that it is not

worthwhile to space the learning of complex tasks. Their analysis contained studies employing a variety of tasks and intervals ranging from a few seconds to 24 hours, but their analysis was heavily weighted toward studies that used short spacing intervals. Since their review was published, a number of new studies on skills that use long spacing intervals have been published, it is therefore interesting to consider whether their findings apply to skill-related studies with spacing intervals of 24 hours or more. Arthur Jr et al. (2010) addressed this issue by training participants in a complex simulation game where participants played the role of the commander of a navy fleet. Participants completed their sessions across 2 weeks or concentrated in 1 week. The 2-week group displayed better learning on a post-test at the end of training than the 1 week group (d = .24) and better retention on a test 8 weeks after training (d = .46). It is important to note that Donovan and Radosevich's (1999) massed groups completed their training all in a single day; if Arthur Jr et al. (2010) had used a similar comparison group they probably would have reported even larger effect sizes.

Additionally, a number of studies have reported the benefits of spacing while learning surgical skills (De Win et al., 2013; Gallagher, Jordan-Black, & O'Sullivan, 2012; Kang et al., 2015; Moulton et al., 2006; Spruit et al., 2014; Verdaasdonk et al., 2007). Some of these studies provide information on effect sizes. These studies generally find a medium to large effect size of spacing and therefore provide additional evidence that spaced repetitions produce a worthwhile improvement in complex skill tasks.

1.3.1.2 Infants and children

In addition to the literature on adults, there are also several studies that have explored the spacing effect for skill-related tasks in infants and children. For example, Vander Linde, Morrongiello, and Rovee-Collier (1985) compared infants who learned to kick to activate an overhead crib mobile with spaced or massed practice. The daily group completed three sessions across three consecutive days,

the alternate-day group completed three sessions on alternate days, and the massed group completed all three sessions on a single day. Consistent with the adult literature, infants in the alternate-day group learned the task significantly faster than infants in the daily and massed groups.

Studies have also investigated the spacing effect for retention in infants, utilising the concept of a time window (Rovee-Collier, 1995). The time window is a limited period of time in which additional experiences can be integrated into a memory, beyond which the time window shuts because the memory has been forgotten. Rovee-Collier (1995) compared time windows to critical periods in that time windows are a limited period of time in which an organism is responsive to certain experiences, except that the time window is psychological rather than biological. For example, in one study, a new item was integrated into a pre-existing category if it occurred 4 days after the original category learning experience, but the new item was treated as a unique event if it occurred 5 or 6 days after category learning (Rovee-Collier, 1995). The time window concept uses the same basic principles to explain a range of phenomena in infant memory which involve integrating new experiences with related long-term memories (Rovee-Collier, 1995). Some of the phenomena explained using the time window concept are categorization, memory modification, and the spacing effect (Rovee-Collier, 1995). For the spacing effect, studies using the time window concept have found that repetitions that are later in the time window lead to a task being remembered longer than repetitions that are earlier in the time window; however, if the repetition is outside the time window, even if it is only a single day, it is as if the infant is encountering the stimulus for the first time (Galluccio & Rovee-Collier, 2006; Hartshorn, Wilk, Muller, & Rovee-Collier, 1998; Hudson & Sheffield, 1998; Rovee-Collier, Evancio, & Earley, 1995).

Rovee-Collier et al. (1995) illustrated the time window concept by employing a crib mobile paradigm similar to the one described above. In this paradigm 3-month-old infants completed two 15 minute sessions with a space of 1, 2, 3 or 4 days between sessions. Retention was then tested 8 days after the first session and the results are illustrated in Figure 5. Infants whose second session occurred

2 days after the first session had perfect recall in the retention test because their second session was late but still within the infants' time windows. In contrast, infants whose second session was 4 days after their first returned to their baseline level of performance because the second session occurred outside of their time window. Finally, infants whose second session was 1 day or 3 days after the first session showed intermediate retention. For the 1 day space this was because the second session was early in the time window, whereas for the 3 day space it was because some infants could retrieve the memory whereas others could not (i.e., outside of the time window for some infants but just within the time window for others). The infants who failed to retrieve the memory did not benefit from the second session and performed close to baseline at the retention session, while those who retrieved the memory did benefit from the second session, suggesting that the optimal space is just before a memory is forgotten. Overall, these results show that there is an inverted-U relationship for spacing and retention in infants.

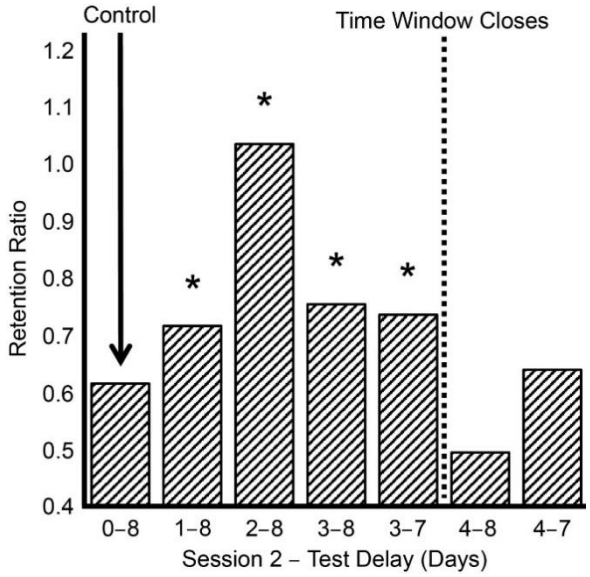

Figure 5. Retention performance of infants with varying delays between session 1 and session 2 and session 2 and the test from Rovee-collier et al. (1995). *Significantly different from baseline performance.

Interestingly, another study has suggested that the finding that the optimal space is just before the memory is forgotten may not be the full story. Hsu (2010) examined how long 6-, 9-, 12-, 15- and 18-month old infants retained a memory for an operant task, equivalent to the mobile paradigm, when their second session was completed near the end of their time window. Comparing her data with an earlier study that used the same methodology and completed the second session 24 hours after the first (Hartshorn, Rovee-Collier, et al., 1998), Hsu (2010) concluded that for 6 month old infants completing the second session near the end of the time window resulted in better retention, but for the 9- to 18-month old infants completing the second session near the end of the time window lead to worse retention than a 24 hour space. It is important to note that the 9- to 18-month old infants successfully retrieved the memory in the second session; thus if Hsu's (2010) conclusions are correct this means the assumption that more difficult retrievals are always better, as suggested by some accounts of the spacing effect (e.g., Bjork & Bjork, 1992; Delaney et al., 2010), may not be correct. However, given the use of a between-study comparison, it would be desirable for Hsu's (2010) finding to be replicated.

1.3.2 Language and Verbal Tasks

1.3.2.1 Adults

Unlike skill-related tasks, for language tasks spacing leads to equal or worse learning but enhanced retention. While the finding of no spacing effect for learning in language-related tasks may seem unusual, a close reading of the studies referenced reveals that this finding is very consistent (Bahrick, 1979; Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bloom & Shuell, 1981; Cepeda et al., 2009; Cepeda et al., 2006; Keppel, 1964, 1967; Moss, 1996; Simone, Bell, & Cepeda, 2013; Suzuki & DeKeyser, 2015). For example, Bloom and Shuell (1981) compared a spaced group that studied French words on three consecutive days to a massed group that studied French words all in the same day. On an immediate test, conducted to assess learning, the spaced and massed groups were equivalent, but on a retention test 4 days later the spaced group's performance was superior to the massed group, thus

spacing led to equivalent learning but enhanced retention. In contrast, for a skill-related task, Shea et al. (2000) compared a spaced group that practiced a discrete timing task on three consecutive days to a massed group that completed the same amount of practice within the same day and found that the spaced group performed better at the end of training. The possible reasons for spacing enhancing learning in skill-related tasks but not language-related tasks will be discussed in later sections.

For retention, similar to Rovee-Collier et al.'s (1995) findings with infants, an inverted-U curve for the spacing effect has been reported. For example, Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) had participants study 32 facts across 2 sessions and then conducted a retention test. In the first session the facts were studied and tested until each fact was correctly recalled and in the second session facts were tested twice with feedback. The spacing interval between the first and second session varied across participants, ranging from 0 to 105 days. Similarly the delay between the second session and retention test ranged between 7 and 350 days. Initially, retention improved as the spacing interval increased but then declined forming an inverted-U curve (Figure 6). Additionally, the optimal space varied depending on the retention delay with the optimal space being longer for longer retention delays (e.g., for the 7 day retention delay the optimal space was 3 days and for the 35 day retention delay the optimal space was 8 days).

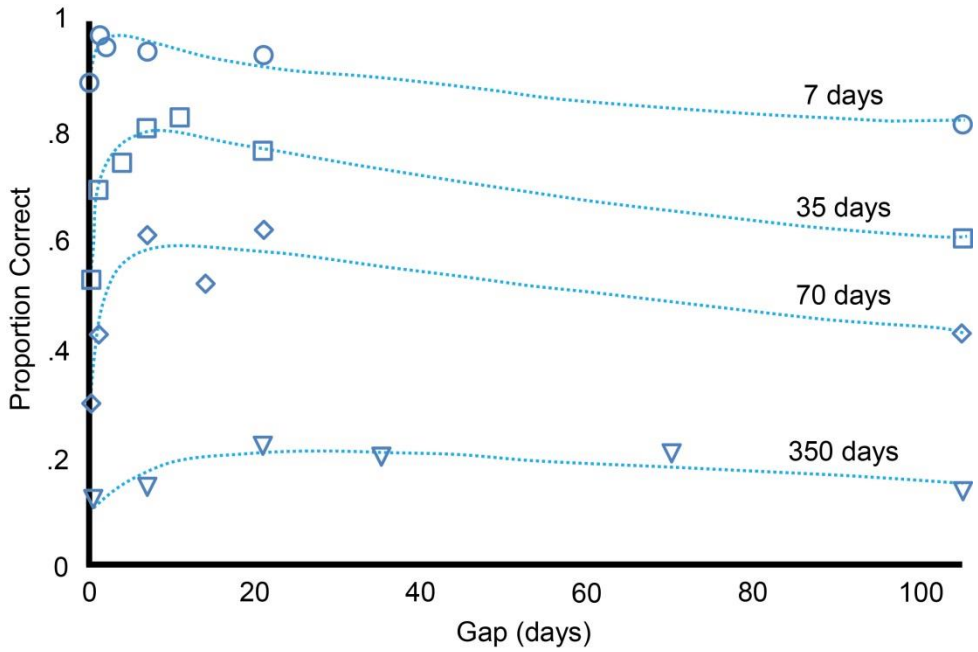

Figure 6. Retention performance of participants who repeated facts with varying spacing and retention test delays. Figure redrawn from Cepeda et al. (2008).

Other studies have not used as many spacing and retention delays as Cepeda et al. (2008), but the finding that the optimal spacing interval changes depending on the retention delay has been reported for re-reading texts (Rawson, 2012; Rawson & Kintsch, 2005; Verkoeijen, Rikers, & Özsoy, 2008), word-pairs (Bahrick & Phelphs, 1987; Gerbier, Toppino, & Koenig, 2015; Küpper-Tetzel & Erdfelder, 2012; Küpper-Tetzel, Kapler, & Wiseheart, 2014) and for remembering vocabulary (Küpper-Tetzel, Erdfelder, & Dickhäuser, 2014). One interesting question yet to be directly addressed is what effect increasing the number of re-learning sessions has on the inverted-U curve. From the studies published to date, I hypothesise that as the number of sessions increases, the number of spacing intervals that could be considered optimal or close to optimal for a particular retention delay increases, reflecting a widening of the inverted-U curve. For example, Küpper-Tetzel, Erdfelder, et al. (2014) used 1 relearning session and found that for a 7-day delay a 1-day spacing interval produced recall of about 86%, whereas a 10-day space produced recall of around 62%. In contrast, Bird (2010) used 4 relearning sessions and found for a 7-day delay a 3-day spacing interval produced retention of 83.1% and a 14-day space produced retention of 80.9% (i.e., not significantly different from the 3-day-spacing group).

1.3.2.2 Infants and children

While adults do not learn more with spaced presentations in language tasks, children do seem to learn more from spaced presentations. For example, Ambridge, Theakston, Lieven, and Tomasello (2006) exposed 4 year old children to 10 sentences containing a grammatical construction they had not yet learned. The exposures were massed all in a single session or spaced across five consecutive days. Children in the spaced group showed much better learning than those in the massed group on a test immediately after their last training trial. In contrast, Miles (2014) had adult Korean students learn English grammar and massed their learning into one day or spaced their learning across multiple days separated by varying delays and found that on the immediate test the spaced and massed groups learning was approximately equal. Thus, using a very similar design, spacing seems to enhance learning in children but not in adults. Another way of presenting this finding is that, at very short retention intervals, massing is better than or equal to spacing for adults but not for children (Cepeda et al., 2006; Maddox, 2016).

Additionally, in children's language tasks manipulating the intensity of training parallels the findings for adults learning skills. Schwartz and Terrell (1983) found that children learned more words when presentations were spread over 10 days compared to 5 days. Similarly, Childers and Tomasello (2002) found that children learned to produce more words when presentations were distributed over 4 consecutive days rather than 2 consecutive days. Childers and Tomasello (2002) also found that when children's learning occurred across 4 sessions there was no difference in learning between groups who had an intersession interval of 24 hours or 3 days, somewhat contradicting the spacing effect. There are several possible explanations for this finding, the first is that the absolute time between repetitions is not particularly important, but it is necessary for the child to have a period of sleep between each session before additional learning can occur. The second explanation is that the 3 day interval does

enhance learning, but its benefits are undermined by greater forgetting which leads to performance equivalent to the 24 hour intersession interval.

Other studies have found spacing enhances children's retention in language-related tasks (Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2012; Moinzadeh, Talebinezhad, & Behazin, 2008; Sobel, Cepeda, & Kapler, 2011). One study conducted with children is particularly interesting for its practical implications. Moinzadeh et al. (2008) compared five groups of 12 – 13 year olds learning English as a foreign language. All of the groups completed six sessions: one group completed two sessions per day, a second group completed one session per day, a third group completed one session every alternate day, a fourth group completed two sessions per week and a fifth group completed one session per week. Learning was assessed via a test conducted one day after the final learning session and retention was assessed via a test conducted one month after the final learning session. Moinzadeh et al. (2008) reported that the group with one session per day performed the best on the learning test and the group with one session every alternate day performed the best on the retention test. This suggests that when considering the optimal spacing schedule you should consider how regularly the language will be used. For example, suppose a 12 year old is moving to a foreign country permanently, this study suggests that it would be optimal for them to learn with daily sessions; if however, the child was going to a foreign country for a holiday and most likely would return for multiple holidays across their lifetime, then it would be optimal for them to learn with sessions on alternate days.

1.3.3 Puzzles and Maze learning

Interestingly, the influence of spacing seems to be different when learning mazes and puzzles. For spacing intervals of 24 hours, spacing is more advantageous for mazes with a greater number of choice points (i.e., a larger maze pattern). Pechstein (1917) had participants learn a covered stylus maze that had to be navigated using participants sense of touch and movement. When participants learned the maze as a whole spaced practice was superior to massed practice. However, when participants

learned the maze by different methods which involved dividing the maze into parts and learning these parts individually, the advantage of spaced practice diminished and in some cases the massed group performed better than the spaced group. Cook (1937) obtained a similar result with a more direct manipulation. Cook (1937) had participants learn finger mazes with 4, 8, or 12 choice points. Cook (1937) found that overall the massed group had fewer errors than the spaced group but that the superiority of the massed group diminished as the number of choice points increased. At 4 choice points the massed group had a superiority of 98%, at 8 choice points the superiority was 28% and at 12 it was 10%. Similarly, across the different choice points the spaced group completed the maze in less time than the massed group but this advantage increased as the number of choice points increased. For example, at 4 choice points the spaced groups superiority was 4%, at 8 it was 13% and at 12 it was 20%.

For learning to solve a puzzle massed practice generally seems to be better than spaced practice. Cook (1934) had participant's complete 20 trials within a single day (massed) or through completing 1 trial per day (spaced). The participant's task was to arrange the puzzle pieces into the correct configuration (a t or a cross depending on the puzzle). Cook (1934) found that massed practice led to faster solution times than spaced practice, however, spaced practice lead to better retention. Another study by Ericksen (1942) had participants learn to solve a puzzle box or a 10-unit maze with a similar pattern of spacing or massing. Ericksen (1942) found that massed practice lead to better learning of the puzzle box but worse learning of the maze compared to spaced practice. Finally, Garrett (1940) compared the learning of three different "mental" puzzle tasks. The tasks were a code-learning task, a symbol-digit substitution task, and an artificial language task. Based on data he collected from his participants Garrett determined that the code-learning and symbol-digit substitution tasks were less complex than the artificial language task. Furthermore, he found that the code-learning and digit

substitution tasks were better learned under spaced practice and the artificial language task was better learned under massed practice.

What are we to make of these findings? One factor that may be involved in the pattern of results is the amount of forgetting produced by each task. It may be the case that the precise steps required to solve the puzzle tasks are forgotten more quickly than the choices in the maze tasks where there may be a set of contextual cues to remind participants about the particular direction. This influence of forgetting is likely reflected in Garrett's (1940) study where the artificial language task had a more complex set of rules to remember than the symbol-digit substitution and the code-learning tasks. It is also possible to observe a similar effect of complexity in mazes. A study by Cook (1944) manipulated the number of choices at each choice point in the maze (instead of the total number of choice points like Cook, 1937) and found that a greater number of choices at each point of the maze favoured massed practice presumably because it was then harder to remember which route to go down. This explanation, however, leaves unexplained why a larger maze pattern (i.e. a greater total number of choice points) favours spaced practice. Additionally, since these studies are quite old and no direct replications have been conducted, it would be worthwhile for someone to conduct replications of these studies to confirm the pattern of results observed.

1.3.4 Generalization

1.3.4.1 Adults

Interestingly, a few studies have found that spacing not only benefits the learning and retention of specific items but improves the generalization of learning. Hagman (1980) had participants learn and practice electrical testing on the same equipment or different equipment, with practice massed all in one day or spaced on three consecutive days. On a transfer test after a 2-week delay, spaced practice on different equipment resulted in better transfer than spaced practice on the same equipment. Spaced practice on the same equipment resulted in better performance on the transfer test than massed practice

on the same or different equipment. Moreover, massed practice on the same or different equipment resulted in equivalent performance on the transfer test, indicating that spacing was necessary for training variations to promote generalization. Similarly, Moulton et al. (2006) compared massed and spaced groups who practiced microsurgical skills on PVC-artery models and arteries from a turkey thigh, and tested to what extent their skills transferred to a live rat 1 month after training. Moulton et al. (2006) found that the spaced group performed significantly better on a variety of outcome measures than the massed group. There is one other study that claims to show transfer for diagnostic skills but, because it used a within-subjects design and spaced training was always completed before massed training, it is not possible to know whether this was due to experience alone or the spacing effect (Kerfoot et al., 2010).

The two studies described above are the only studies I am aware of that systematically examine the effect of long spacing intervals on generalization in adults, obviously more research is needed on this aspect of the spacing effect. However, based on these studies and the studies on children discussed below it seems probable that the spacing effect will enhance generalization in other circumstances in adults.

1.3.4.2 Infants and Children

Studies with children have investigated the impact of spacing on generalization using a greater range of spacing intervals relative to the adult literature. For example, Vlach and Sandhofer (2012) investigated the impact of spacing on the generalization of simple and complex science concepts in 5 to 7 year olds. The children in their study completed four lessons on biomes, with each lesson involving a different context (desert, grasslands, artic, ocean or swamp), and a post-test 1 week after the last lesson. The massed group completed all 4 lessons in one day, the intermediate group completed 2 lessons per day for 2 days, and the spaced group completed 1 lesson per day for 4 days. For simple

generalization, the spaced group showed significantly greater improvement from the pre- to post-test than the massed group, and the intermediate group's improvement was not significantly different when compared to the massed or spaced groups. In contrast, for complex generalization the spaced group's improvement was significantly better than both the massed and intermediate groups. In fact, the data suggest that the spaced group is the only group to show an improvement in their gain scores as the questions moved from simple to complex, though unfortunately this trend was not tested for statistical significance. Spacing therefore may provide a greater benefit for more complex generalizations.

Gluckman, Vlach, and Sandhofer (2014) replicated Vlach and Sandhofer's (2012) findings but in the post-test, they included questions on the children's memory for facts and concepts talked about during the lessons (e.g., what is a biome?), in addition to generalization questions. The spaced group showed significantly greater improvement than the massed group for simple and complex generalization questions and for memory questions. The reported means displayed the same pattern as above with spacing benefiting complex generalizations more than simple generalizations. Gluckman et al. (2014) also tested for correlations between the memory test and generalization, hypothesising that there would be a positive correlation between memory scores and generalization scores if generalization was related to memory for the lessons. They found no significant correlations, suggesting that in this task memory and generalization may be independent learning processes. Consistent with this finding, Wang, Zhou, and Shah (2014) trained children on working memory games and found no effect of spacing on game performance, but they did find that spacing improved transfer performance on Raven's Progressive Matrices test.

In contrast to the impact of spacing on learning and retention, there has been relatively little research exploring the impact of spacing over long intervals on generalization. However, the studies conducted to date allow us to tentatively conclude that greater spacing (e.g., spreading learning across 4 days vs. 2 days) seems to provide a larger benefit to generalization and that more complex generalizations seem to benefit more from spacing, independent of other more specific forms of

learning. Since generalization and transfer are a very valuable part of learning it would be worthwhile for future research to examine whether these tentative conclusions are reliable and to examine the extent spacing promotes generalization for a greater variety of tasks.

1.4 Summary

Since the majority of studies have been conducted on the spacing effect over short timescales, most of the past literature has ignored the ramifications of the spacing effect over long timescales. When the spacing effect across long timescales is observed separately, some interesting patterns in the literature emerge. The spacing effect occurs for complex tasks over long timescales but does not occur for complex tasks across short timescales. Moreover, spaced repetitions enhance learning in children and adults in skill tasks but do not enhance learning for adults in verbal tasks. Spaced repetitions do, however, seem to enhance retention in both adults and children across different types of tasks over long timescales. Additionally, the spacing effect seems to manifest slightly differently depending of the type of task and this potentially has something to do with participant's prior experience and the complexity of the task.

1.5 The spacing effect in non-human species

A number of studies have also observed the spacing effect over short and long timescales in non-human species. A large number of studies have been conducted using rats in land and water mazes and have found a spacing interval of 24 hours produces better retention than if the trials have been conducted within a single day (Commins, Cunningham, Harvey, & Walsh, 2003; Crystal & Babb, 2008; Goodrick, 1973; McGaugh & Cole, 1965; Morris & Doyle, 1985; Sisti, Glass, & Shors, 2007; Spreng, Rossier, & Schenk, 2002). There are fewer studies that have been conducted across shorter timescales for mazes in rats but the ones that have been conducted have observed a spacing effect

(Kraemer & Randall, 1995; Rick, Murphy, Ivy, & Milgram, 1996). However, greater controls in future research would be beneficial to ensure the effect over short timescales is not due to fatigue (Sisti et al., 2007). Additionally, the spacing effect has been observed over long intervals for other mammalian species such as dogs and horses (Demant, Ladewig, Balsby, & Dabelsteen, 2011; Kusunose & Yamanobe, 2002; Meyer & Ladewig, 2008; Rubin, Oppegard, & Hintz, 1980).

Another set of studies that is of interest is conditioning studies in animals. In these studies the inter-trial interval (ITI) between conditioning trials is varied and longer ITI's lead to a higher rate of acquisition (Bouton & Sunsay, 2003; Detert, Kampa, & Moyer Jr, 2008; Sissons & Miller, 2009). However, this observation is unlikely to be the same as the spacing effect observed in human or in other tasks in non-human species. Unlike the spacing interval in humans, the ITI seems to influence the information or signalling value of the conditioned stimulus (Bouton & Sunsay, 2003; Ward, Gallistel, & Balsam, 2013). In particular, animals with longer ITI's spend their ITI in the conditioning context and this extra exposure means the conditioned stimulus (CS) becomes a better signal for the unconditioned stimulus (US) and the animal learns to a greater degree that the context itself is an unreliable indicator of when the US will occur. This perspective is supported by studies where long ITI rats are taken out of the conditioning context during their ITI and studies where the short ITI rats are given extra exposure to the conditioning context. Both of these interventions abolish the differences in the acquisition of the conditioned response of the short and long ITI groups (Sunsay & Bouton, 2008; Yin, Barnet, & Miller, 1994). In contrast, the spacing effect observed in rats with mazes and tasks in humans and other species seems to be a different phenomenon as generally the spacing interval is spent outside of the learning context (Bahrick & Phelphs, 1987; Demant et al., 2011; Goodrick, 1973; Spreng et al., 2002).

An interesting study of the spacing effect was conducted in honey bees by Menzel, Manz, Menzel, and Greggers (2001). The bees in this study were conditioned by presenting an odour as the conditioned stimulus and sucrose as the unconditioned stimulus. The spaced group had an ITI of 10

minutes and the massed group had an ITI of 10 seconds. Additionally, Menzel et al. (2001) gave some bees a protein synthesis inhibitor before training. Menzel et al. (2001) found that at the end of training the spaced group showed a greater conditioned response than the massed group and that a protein synthesis inhibitor did not interfere with the bees learning of the task but did influence the bees' retention. In particular, when tested 1-2 days later the spaced bees had impaired memory whereas the massed bees did not and when tested 3-4 days later both spaced and massed bees had impaired memory.

Menzel et al.'s (2001) study may be explained in two ways. One interpretation is that the results reflect the distinction between the ITI effect and spacing effect. The ITI effect reflects the rate of responding based on the how informative the animal perceives the CS and is unaffected by protein synthesis. A similar spacing effect to that observed in humans is reflected in the bees' retention of the conditioned response and this is influenced by protein synthesis inhibitors. Additionally, since the bees have a shorter lifespan and a different time-course for consolidation the 10 minute spacing interval in bees might be more like spacing intervals of hours in humans. A second interpretation of this study is that the longer ITI of the spaced group produces a greater response and the maintenance of this stronger response depends on protein synthesis. Although more research is needed to make a precise determination it seems more likely that Menzel et al. (2001) observed two distinct effects based on a study conducted in parasitic wasps (Smid et al., 2007). In this study, parasitic wasps were given three spaced (ten minutes between trials) or three massed Pavlovian conditioning trials where they were able to inject a caterpillar with eggs. Wasps who received spaced trials showed memory performance above naïve wasps when tested 1 – 5 days later whereas wasps who received massed trials did not. Additionally, the long-term memory was not formed if the wasps received a protein synthesis inhibitor before training. Importantly, during the spaced trials the wasps were taken out of the conditioning context and when their memory was tested 1 hour after their trials, massed and spaced wasps showed

the same level of learning, suggesting that the retention advantage Menzel et al. observed for spaced bees may be independent of the level of conditioning.

Based on the evidence it seems likely there is a similar spacing effect in humans, other mammalian species and insects. When analysing their results and referring to past research, researchers should be careful to distinguish between the spacing effect and the ITI effect as these phenomena are likely to be produced by different mechanisms. However, it does seem likely that the spacing effect and the ITI effect can both exist amongst insects on the same task. The spacing effect therefore appears to depend on a basic memory mechanism that is conserved across different species.

2 Chapter 2: Past theories of the spacing effect and a new theory of the spacing effect.

2.1 Can existing theories account for the spacing effect over long timescales?

There have been numerous theories of the spacing effect. Rather than going over many redundant approaches, I will focus on the two predominant types of theories which are still believed to provide a possible explanation for the spacing effect. The first type of theory that is used to explain the spacing effect is contextual variability (Estes, 1955; Glenberg, 1979; Maddox, 2016; Pashler, Cepeda, Lindsey, Vul, & Mozer, 2009; Raaijmakers, 2003). While there are multiple permutations of contextual variability theories, essentially all theories using this mechanism suggest that spaced repetitions lead to a greater variety of contextual elements being integrated into a memory than massed repetitions and a greater variety of contextual elements means that the memory is more likely to be recalled after a delay period. Modern contextual variability theories also have a study-phase retrieval component, whereby the original memory or experience must be recalled during the repetition to integrate additional contextual elements and therefore benefit from the repetition (Maddox, 2016; Pashler et al., 2009; Raaijmakers, 2003). Furthermore, these theories explain the inverted-U curve of the spacing effect by suggesting that recall is based on the match between the test context and the contextual elements integrated during the first presentation and repetitions.

Contextual variability theories work well for explaining the data in verbal studies in adults. Test performance is based on the overlap between contextual elements stored in the memory and the contextual elements present during the test. Performance on an immediate test is often better for massed repetitions because the contextual elements present for the test will be very similar to contextual

elements integrated during the initial presentation and the repetition, leading to a large overlap (Delaney et al., 2010; Maddox, 2016). In contrast, for spaced repetitions the contextual elements present for an immediate test will be similar to the repetition but quite different to the first presentation. On a delayed test the contextual elements will be different to the contextual elements of the first presentation and repetition. It is therefore valuable to have a variety of contextual elements integrated into the memory so there is sufficient overlap, in this case spaced repetitions lead to a greater variety of contextual elements being integrated into the memory than massed repetitions and thus produce better retention (Delaney et al., 2010; Maddox, 2016).

However, when we look outside of the verbal data in adults, contextual variability theories have problems explaining the data especially for the observations of the spacing effect across long intervals in the previous chapter. For long spacing intervals, I found that spaced repetitions led to better performance on an immediate test than massed repetitions on verbal learning tasks in young children and skill tasks in adults. According to the contextual variability theories the reverse should be observed, for an immediate test massed repetitions should lead to equal or better test performance than spaced repetitions due to massed repetitions resulting in a greater overlap between the test's contextual elements and the contextual elements stored as part of the memory. Also difficult to explain with the contextual variability account is why, when using short spacing intervals, the spacing effect is weak or non-existent for complex tasks but robust for complex tasks in which long spacing intervals are used.

A second major class of theories explain the spacing effect in terms of retrieval difficulty (Benjamin & Tullis, 2010; Bjork & Bjork, 1992). In particular, these approaches suggest that greater forgetting occurs for spaced repetitions and this makes retrieval more difficult leading to a greater enhancement in the memory. Retrieval difficulty theories are supported by a number of studies of verbal memory using short spacing intervals (Benjamin & Tullis, 2010; Bjork & Allen, 1970; Crowder, 1976; Hintzman, 1974). Consistent with retrieval being more difficult, many studies of the spacing effect in language-related tasks observe that at the time of the repetition performance is worse for the

spaced group than the massed group (Bjork & Allen, 1970; Cepeda et al., 2009; Glenberg, 1976). In contrast, in many of the skill-related studies using long spacing intervals the opposite is observed, at the time of the repetition performance is better for the spaced group than the massed group suggesting that retrieval of spaced repetitions is easier than retrieval of massed repetitions (e.g., Dail & Christina, 2004; Molloy et al., 2012; Shea et al., 2000). Moreover, a couple of studies have compared the retention performance of Swahili word-pairs with a spaced group that slept during their spacing interval to a spaced group that remained awake during their spacing interval and found that the sleep group performed better at the repetition and subsequently showed better performance on the retention test (Bell, Kawadri, Simone, & Wiseheart, 2014; Mazza et al., 2016). These studies suggest that retrieval difficulty theories may not be able to account for the spacing effect over long timescales.

Another finding noted above was that the spacing effect occurred for perceptual discrimination tasks (Molloy et al., 2012; Wright & Sabin, 2007). Since the discrimination response for each trial is based on stimuli presented in close succession it seems unlikely that retrieval difficulty is influenced by the spacing of repetitions as it potentially is in verbal tasks. Similarly, it seems unlikely that stored contextual elements play a significant role in the ability to make discriminations in these tasks. Significantly, however, for the theory proposed below memory consolidation plays a critical role in improving participant's discrimination skill (Atienza, Cantero, & Stickgold, 2004; Gaab, Paetzold, Becker, Walker, & Schlaug, 2004; Gais, Plihal, Wagner, & Born, 2000; McNaughton, Douglas, & Goddard, 1991).

2.2 A Reconsolidation Account of the Spacing Effect

2.2.1 Reconsolidation as a mechanism

In the past there have been attempts to explain the spacing effect in terms of memory consolidation (Landauer, 1969; Wickelgren, 1972). However, these theories were generally rejected because of theoretical and empirical issues (Bjork & Allen, 1970; Delaney et al., 2010; Dempster, 1989; Hintzman, 1974). In the decades since these papers were published there have been many significant developments in our understanding of consolidation and these developments are what make a reconsolidation account a viable hypothesis for explaining the spacing effect over long timescales.

When the earlier consolidation theories of the spacing effect were published the concept of memory reconsolidation was not widely studied or adopted (Nader & Hardt, 2009; Sara, 2010). Instead the dominant perspective was that a memory was initially unstable and then overtime consolidated in a linear manner (Nader & Hardt, 2009; Sara, 2010). A resurgence of interest in memory reconsolidation lead to experiments showing that this perspective was partially incorrect (Nader & Hardt, 2009); after the initial consolidation period when a memory was retrieved it returned to being unstable and sensitive to disruption. This period of instability probably provides a net benefit as it is necessary for additional experiences to modify and build on the pre-existing memory trace (Alberini, 2011).

One of the ways researchers gained a better understanding of consolidation and reconsolidation was through experiments that used protein synthesis inhibitors such as anisomycin (Nader & Hardt, 2009; Nader, Schafe, & Le Doux, 2000; Suzuki et al., 2004; Wang, de Oliveira Alvares, & Nader, 2009). The initial consolidation experiments injected a protein synthesis inhibitor a little before or after training and found that memory was generally unaffected 0 to 2 hours after training but when the memory was tested 24 hours after training it was disrupted (Davis & Squire, 1984; Goelet, Castellucci, Schacher, & Kandel, 1986; McGaugh, 2000; Meiri & Rosenblum, 1998; Schafe, Nadel, Sullivan, Harris, & LeDoux, 1999). Thus a short-term memory could be sustained for a few hours without

generating new proteins but new proteins were needed for a long-term memory. Later, when research on reconsolidation developed, similar findings were observed (Nader et al., 2000; Rossato et al., 2007; Schafe & LeDoux, 2000; Suzuki et al., 2004). Researchers discovered that when the memory was retrieved 24 hours or more after training, injecting protein synthesis inhibitors into brain areas associated with the memory disrupted the memory after 24 hours but not when tested a few hours after training. These findings provide several important pieces of information about consolidation and reconsolidation. Firstly, the neural consolidation processes which influence the development of the long-term memory take time to develop and may not affect the memory over the first few hours after the initial training or a reactivation. Secondly, the memory gets additional consolidation (reconsolidation) from the reactivation.

Other experiments using protein synthesis inhibitors and a variety of other techniques have further developed our understanding of memory reconsolidation. Researchers have determined two functions for memory reconsolidation: altering an existing memory trace in response to new experiences and strengthening a memory (Alberini, 2011; Alvares et al., 2013; Inda, Muravieva, & Alberini, 2011; Lee, 2008). On a behavioural level, memory strengthening has been identified as improved learning and better retention (Inda et al., 2011; Lee, 2008; Morris et al., 2006). Additionally, research has shown that memory reconsolidation seems to be a basic memory process occurring across many different tasks and species (Alberini, 2005; Forcato et al., 2007; Pedreira, Pérez-Cuesta, & Maldonado, 2004; Walker, Brakefield, Hobson, & Stickgold, 2003). In particular, it has been demonstrated in humans using both motor skill and verbal tasks (Coccoz, Maldonado, & Delorenzi, 2011; De Beukelaar, Woolley, & Wenderoth, 2014; Forcato, Argibay, Pedreira, & Maldonado, 2009; Forcato et al., 2007; Walker et al., 2003). Based on this research we can be confident that the reconsolidation process is playing a role in the experiments described earlier. However, the critical question is: does the time between repetitions influence the degree to which consolidation and

reconsolidation strengthen and improve memories? Or alternatively, is reconsolidation's effect on memory independent of the timing of repetitions and merely mediates another mechanism which is responsible for the spacing effect? For example, it might be the case that the reconsolidation process is responsible for integrating additional contextual elements into a memory and it is the addition of these elements which produce the spacing effect. Before I answer this question directly, it is worthwhile to address another development in our understanding of the consolidation of memory.

A second development in memory consolidation research is a much better understanding of the significance of sleep (Stickgold, 2006). Sleep plays an active role in memory consolidation. During sleep memories are reactivated (Ji & Wilson, 2007; Oudiette & Paller, 2013; Pavlides & Winson, 1989; Wilson & McNaughton, 1994) and the different stages of sleep are associated with different tasks and aspects of performance, suggesting that sleep-based consolidation makes a qualitatively different contribution to memory than the waking state (Gais, Mölle, Helms, & Born, 2002; Marshall, Helgadóttir, Mölle, & Born, 2006; Stickgold, 2006; Walker & Stickgold, 2004). On a behavioural level there are parallels between the sleep literature and the spacing effect studies I reviewed earlier. In skill learning tasks a period of sleep leads to better performance with no additional practice (Fischer & Born, 2009; Kuriyama, Stickgold, & Walker, 2004; Walker et al., 2003) and in verbal tasks sleep generally reduces forgetting but does not improve performance (Abel & Bäuml, 2014; Drosopoulos, Schulze, Fischer, & Born, 2007; Lahl, Wispel, Willigens, & Pietrowsky, 2008). Additionally, sleep is important for the generalization of memories (Stickgold & Walker, 2013). Similar to the reconsolidation literature, studies investigating sleep and learning indicate that consolidation during sleep is influencing the same dependent variables (learning and retention) as the spacing effect and sleep (like the spacing effect) requires time to influence these variables. Thus, we return to the question does the spacing of repetitions influence the benefit gained by sleep-consolidation and reconsolidation? Or are the benefits from sleep-consolidation and reconsolidation independent of the spacing of repetitions?

Logically, it seems likely that the spacing of repetitions would influence the consolidation and reconsolidation processes and their beneficial effects on learning and retention. There are multiple studies indicating that consolidation during the night is influenced by an individual's learning experiences during the day (Laureys et al., 2001; Maquet et al., 2000; Pavlides & Winson, 1989; Poe, Nitz, McNaughton, & Barnes, 2000). For example, Gais et al. (2002) found that participants who learned word-pairs showed a greater density of sleep spindles on the following night than participants who completed a word task that did not require long-term memory. It seems likely, therefore, that the spacing of repetitions over different numbers of days might influence memory consolidation during sleep. For example, in the Arthur Jr et al. (2010) study reviewed earlier the spaced group learned the naval command simulation over two weeks and the massed group learned it over 1 week. It seems probable that the spaced group might get a greater degree and quality of reprocessing during sleep than the massed group.

There is also some experimental evidence that suggests that the spacing of repetitions influences consolidation and reconsolidation. Before we look at this evidence, it is worthwhile to state more clearly what the reconsolidation account of the spacing effect entails. Essentially, I am suggesting that greater time between repetitions provides more time for the memory to consolidate and this greater degree of consolidation makes the additional consolidation (reconsolidation) induced by a repetition more effective at enhancing the memory. Furthermore, part of the reconsolidation process is further processing of the memory during sleep.

Many of the verbal and skill tasks reviewed earlier in this paper used a design where the massed group completed all of their trials in one day and the spaced group completed their trials across two days. One study that investigated memory reconsolidation had a similar set-up. In this study rats completed two trials of context-shock conditioning (Lee, 2008). Some of these rats completed the two trials all in a single day and others completed the trials across two days and the memory of both groups

was tested on the third day. After their second trial the rats were either injected with a substance that inhibited Brain Derived Neurotrophic factor (BDNF), a protein that has a variety of functions related to neuron growth and neural plasticity or ZIF268, a transcription factor involved in learning and memory processes. The researchers found that when the second trial occurred on the first day, inhibiting BDNF expression disrupted the memory but inhibiting ZIF268 expression had no effect. In contrast, when the second trial occurred on the second day inhibiting BDNF had no effect but inhibiting the expression of ZIF268 disrupted the memory. Furthermore, if no reactivation trial occurred on the second day inhibiting ZIF268 had no effect on memory performance. The findings of this study suggest that a repetition engages slightly different neural mechanisms depending on the spacing of the repetition. In particular, my interpretation of this study is that ZIF268 is particularly important only on day 2 because it is part of the process building on the consolidation that occurred the previous night. Consistent with this perspective, other researchers have also found that some of the mechanisms used by memory reconsolidation are different to the initial consolidation process (Bahar, Dorfman, & Dudai, 2004; Bozon, Davis, & Laroche, 2003; Taubenfeld, Milekic, Monti, & Alberini, 2001).

In mammals some memories shift from being dependent on the hippocampus to being dependent on the cortex. This process is generally called systems consolidation and is thought to be beneficial for long-term memory (Alvarez & Squire, 1994; Milner, 1989). If spacing repetitions allows more time for consolidation and this consolidation makes the subsequent reconsolidation process more effective this should lead to a greater degree of systems consolidation. Lehmann et al. (2009) conducted a study where one group of rats received 12 context-shock trials spread across 6 days (i.e., spaced rats) and another group received 12 context-shock trials all in one day (i.e., massed rats). The rats then had their hippocampus lesioned 7 to 10 days after their initial learning session. The spaced rats continued to show fear to the context whereas the massed rats were amnesiac, demonstrating that the spacing of repetitions increased the rate of systems consolidation such that the memory was hippocampal-independent in spaced rats but hippocampal-dependent in massed rats. This study therefore provides

some initial evidence that spaced repetitions enhance the consolidation of memories to a greater extent than massed repetitions.

The relationship between the spacing of repetitions and memory consolidation was also explored in a study by Vilberg and Davachi (2013) using a within-subjects design. Participants studied and restudied word-object pairs and word-scene pairs. Massed pairs were restudied after 20 minutes, while spaced pairs were restudied after 24 hours, and memory for all of the pairs was tested 24 hours after the restudy period. While the participants were restudying the pairs they were scanned in an fMRI. Spaced word-object pairs remembered on the test showed a greater connectivity between the hippocampus and the perirhinal cortex than massed word-object pairs that were remembered. Additionally, the likelihood of spaced word-object pairs being forgotten could be predicted by the connectivity between the hippocampus and the perirhinal cortex, but the same prediction could not be made for the massed word-object pairs. The results for word-object pairs are consistent with the proposal that allowing time for consolidation enables the spaced repetition to be more effective than the massed repetition. No relationships were identified for the word-scene pairs, but this may have been due to these pairs consolidating differently and the measures used were unable to detect their consolidation.

Additional evidence that reconsolidation is more effective for partially consolidated memories comes from a study by Tse et al. (2007). In their study, rats learned the spatial arrangement of an arena and that a flavoured pellet in the start box meant that food corresponding to that flavour was hidden in a specific sand-well in the arena (there were six flavours and six sand-wells). After the initial task was learned over several weeks and consolidated, Tse et al. (2007) introduced two new sand-wells into the arena that were associated with new flavours. Learning these new flavour-location associations would have induced reconsolidation as the memory for the arena would be modified to integrate the new learning. The two new flavour-location associations were learned after a single trial and were

independent of the hippocampus 48-hours later. Normally, new associations take several weeks to become independent of the hippocampus, suggesting that the framework the rats had established greatly increased the rate of memory consolidation. Additionally, reinforcing their findings, Tse et al. (2011) found that learning new associates within a previously established arena led to a greater expression of genes associated with plasticity in neocortical areas than learning new associates in a new arena or retrieval of previously learned associates. Like the other studies reported on, Tse et al. (2007; 2011) results support the central principle of the reconsolidation account of the spacing effect, that allowing memories time to consolidate enhances the reconsolidation of memories. Tse et al.'s (2007; 2011) studies were designed to gain a better understanding of the schema effect, which is the observation that establishing a framework of knowledge facilitates memory for additional learning that can be fit within the same framework. The spacing effect over long timescales is likely to partially overlap with the schema effect, the core difference being that with the spacing effect, the additional learning has a much higher degree of similarity to the established framework.

From our review of the evidence relating reconsolidation to the spacing effect, we can establish why spaced repetitions might be beneficial for learning and retention. For learning, spacing enables some initial learning to consolidate and then at the repetition the reconsolidation process can more effectively integrate and consolidate additional learning thus building on the earlier consolidation process. For retention, spacing enables the memory to consolidate and then the subsequent reconsolidation process is more effective at enhancing the memory, making it more durable.

2.2.2 Reconsolidation and the inverted-U curve of the spacing effect

I have discussed why a longer spacing interval produces better memory than a shorter interval. However, as discussed earlier, longer spacing intervals are not always better than shorter intervals and there is an inverted-U curve for the spacing interval and its effect on retention. To better understand the inverted-U curve, it is worth comparing the curve for infants and adults. For adults the inverted-U

curve shifted depending on the length of the retention delay, with the optimal space being longer for longer retention delays (Cepeda et al., 2008); for infants (Galluccio & Rovee-Collier, 2006; Rovee-Collier et al., 1995), changing the retention delay for the same set of spaces was not directly tested, but for 3-month old infants going beyond the optimal spacing interval lead to the infant performing at baseline, which means the optimal interval could not have shifted with a longer retention delay. I think that the inverted-U curve for both adults and children can be accounted for by assuming that forgetting reduces the effectiveness of memory reconsolidation.

The data for 3-month old infants' retention of the crib mobile paradigm is relatively simple to explain (Rovee-Collier et al., 1995). The two day spacing interval produced the best retention (see Figure 5, pg. 20) because retrieval is successful in all or almost all of the infants and sufficient time has passed to enable the reconsolidation process to be quite effective. For the 3 day interval some of the children retrieved the memory initiating reconsolidation and others did not, leading to a return to baseline performance on the retention test, the intermediate level of retention is a result of averaging across these two sub-groups. For the 4 day interval the majority of the children can no longer retrieve the memory thus reconsolidation does not occur leading to very poor retention.

The same principles can be used to understand the inverted-U curve observed by Cepeda et al. (2008) (Figure 6, pg. 23). It is important to note that the infant's retention was based on performance on a single task whereas in adult's retention it was based on the recall of 32 facts and this difference most likely leads to the observed differences in the inverted-U curves. It seems a reasonable assumption that amongst the 32 facts there were differences in strength based on factors such as the attention given to that fact during learning or the memorability of the fact. These differences in strength might result in a slightly different forgetting curve for each fact and a different point in time when it is optimal to repeat each fact. The optimal spacing interval for a particular retention delay across all of the facts is the one long enough to provide a substantial benefit for some facts but short enough so that not too

many facts are so weak that they receive little or no benefit from the repetition. The optimal spacing interval and other points on the curve shift depending on the retention delay, because as the retention delay increases the facts need more time before the repetition will be beneficial enough to be recalled after the longer, more difficult delay. However, as a consequence of the repetition occurring later, some weaker facts that would have benefited from an earlier repetition receive little or no benefit from the repetition, but these weaker facts would have been forgotten across the longer retention delay anyway. For example, for a retention delay of 7 days the optimal spacing interval is around 3 days, hypothetically this could mean that 98% of the facts benefit from the repetition and 2% of the facts are too weak to benefit and this results in 94% of the facts being recalled in the retention test. For a retention delay of 35 days, the optimal spacing interval is around 8 days. At this retention delay, a 3 day spacing interval still benefits 98% of the facts, but due to the long delay only 70% of the facts are remembered in the retention test. In contrast, for a spacing interval of 8 days, there is greater forgetting *before* the repetition, perhaps resulting in only 90% of the facts benefiting from the repetition. However, those facts which do benefit from the repetition receive a larger benefit, leading to less forgetting. Therefore at the 7 day retention delay, the 8 day spacing interval might lead to recall of 88%, which is worse than the recall of 94% produced by the 3 day spacing interval; but at the 35 day retention delay, the 8 day spacing interval might result in recall of 80%, which is better than the 70% recall at the 3 day spacing interval. Thus by taking into account that multiple facts are learned and that there are probably differences in the benefits that the facts receive from repetitions, the core characteristics of the inverted-U curve for retention can be accounted for.

2.2.3 Accounting for the effects of spacing on learning

One of the problems with some of the existing theories of the spacing effect above was that they are unable to explain why under some circumstances spacing benefits learning and retention and while in others it benefits retention but does not enhance learning. One finding is that the spacing effect

enhances learning in language-related tasks in children but not in adults. Noticeably, the adults can learn a lot of words or word pairs within a single session. Adults for example can learn 40 new word-pairs in the first session (e.g., Bahrick & Hall, 2005; Cepeda et al., 2009). In contrast, children's ability to learn new words is more limited, with spacing studies teaching children 6 to 16 new words, which are not all remembered even after multiple learning sessions (Childers & Tomasello, 2002; Schwartz & Terrell, 1983). Part of the reason the spacing effect occurs for children and not adults in this context might be that children require time for consolidation between presentations, whereas adult's rapid learning makes consolidation unnecessary for acquisition, but they forget the words relatively quickly and do require consolidation and reconsolidation to make long-lasting memories.

Another finding is that adults show the spacing effect for learning skill-related tasks but not language-related tasks. A similar finding occurs in the sleep literature, after a period of sleep performance improves in skill tasks (De Beukelaar et al., 2014; Fischer & Born, 2009; Kuriyama et al., 2004; Mednick, Nakayama, & Stickgold, 2003; Walker et al., 2003) but performance generally declines after a period of sleep in language tasks with sleep's beneficial effects occurring due to reduced forgetting (Abel & Bäuml, 2014; Drosopoulos et al., 2007; Lahl et al., 2008; Stickgold & Walker, 2007; Yaroush, Sullivan, & Ekstrand, 1971). Since I believe that sleep consolidation contributes to the spacing effect, the findings in the spacing literature and sleep literature can be explained in a similar way. Noticeably, in skill-learning tasks participants are generally only learning one skill and their acquisition is gradual occurring over days. In contrast, for language tasks words or word-pairs are generally acquired rapidly with the difficulty of the task coming from the large number of words they have to learn. The explanation for the difference between language-related and skill-related tasks is thus essentially the same as the explanation for the difference between children and adults for language-related tasks. After a certain number of repetitions within a single day, skill-related tasks require consolidation for additional improvements in performance, whereas in language-related tasks

repetitions within a single day remain effective until the word-pairs are acquired but consolidation and reconsolidation is necessary for other improvements in the memory such as reduced forgetting and resistance to interference. Why might there be differences in acquisition for adults in language tasks compared to skill tasks and language tasks in children? A plausible explanation is adults daily experience with language facilitates acquisition in language tasks and adults do not have the same degree of experience in motor skill tasks and children obviously do not have the same degree of experience with language as adults. An implication of this idea is that an adult's expertise or experience in a particular area will impact the benefits received from the spacing effect. For example, spacing might be less beneficial for expert pianists learning a new piece than for novice pianists.

The observation that the spacing effect enhances the generalization of learning fits well with the reconsolidation account. Numerous studies have demonstrated that one of the important functions of consolidation is the generalization of learning (Fischer, Drosopoulos, Tsen, & Born, 2006; Friedrich, Wilhelm, Born, & Friederici, 2015; Gómez, Bootzin, & Nadel, 2006; Stickgold & Walker, 2013). Therefore, if the reconsolidation induced by spaced repetitions enhances the consolidation processes (i.e., as reported for systems consolidation earlier in the study by Lehmann and McNamara (2011)), you would expect generalization to be enhanced as well. Additionally, the finding that spacing benefits complex generalizations more than simple generalizations is paralleled in the consolidation literature where the more complex parts of a task receive the greatest benefit from memory consolidation (Kuriyama et al., 2004).

2.2.4 The reconsolidation account in comparison to previous consolidation accounts

There are significant differences between the use of reconsolidation as a primary mechanism to explain the spacing effect and the use of consolidation as a mechanism of the spacing effect as explored in earlier consolidation accounts (Landauer, 1969; Wickelgren, 1972). An important difference is the significance placed on memory retrieval. Retrieval of the original memory trace is necessary for the

reconsolidation process which involves a period of instability and allows for modifications of the memory trace (Alberini & LeDoux, 2013; Sara, 2000). Interestingly, the development of retrieval's significance in the reconsolidation literature is essentially the same as the concept of study-phase retrieval which was developed in the spacing effect literature. Study-phase retrieval is the observation that the original experience or memory for an item is often retrieved when a repetition occurs and it has been observed that study-phase retrieval is necessary for the spacing effect to occur (Delaney et al., 2010; Hintzman et al., 1975; Thios & D'Agostino, 1976). Study-phase retrieval is therefore an intrinsic part of a reconsolidation account of the spacing effect while it was not part of earlier consolidation accounts.

The role that retrieval plays in modifying the memory trace leads the reconsolidation account to different predictions than previous consolidation accounts. In Landauer's (1969) consolidation account of the spacing effect, when an item was presented twice, both presentations initiated a consolidation process and memory performance is a summation of the consolidation initiated by these two presentations. However, the consolidation initiated by the second presentation disturbs the consolidation of the first presentation, therefore, a massed repetition leads to less total consolidation and poorer memory performance than a spaced repetition because the consolidation of the first presentation is disturbed soon after it is initiated (Hintzman, 1974; Landauer, 1969). The implication of this theory is that the locus of the spacing effect is on the first presentation of an item rather than the second. Empirical evidence later indicated that the second presentation or retrieval of a memory is more likely to be the locus of the spacing effect (Hintzman, 1974; Hintzman, Block, & Summers, 1973). In a memory reconsolidation account of the spacing effect the locus of the spacing effect is on the second presentation. This part of the theory is derived generally from the fact that in the reconsolidation literature retrieval is acting to modify the memory trace and is supported more specifically by the Lee (2008) study described earlier. Recall, in the Lee (2008) study, when two

context-shock trials occurred in a single day inhibiting BDNF after the second trial disrupted the memory but ZIF268 did not, whereas when the two trials occurred across two days inhibiting BDNF had no effect but inhibiting ZIF268 disrupted the memory. For concision I left out an additional condition Lee (2008) included, whereby the same procedure was followed except that the rat received only one context-shock trial and BDNF or ZIF268 was then inhibited. In this case the results were identical to when two trials occurred on the same day, inhibiting BDNF disrupted the memory while inhibiting ZIF268 had no effect. This suggests that that the effectiveness of the second trial was undermined by having it on the same day as the first trial.

Another difference between the reconsolidation account and the earlier consolidation accounts is the reason why delaying the repetition is beneficial. For example, Landauer (1969) emphasised that delaying the repetition is important because it increases the *amount* of consolidation. In the earlier sections I have already reviewed evidence that the repetition engages different neural processes depending on when it occurs (e.g., see the earlier discussions of Lee, 2008; Tse et al., 2007) and that consolidation during sleep makes important qualitative changes to the memory. The reconsolidation account, therefore, puts greater emphasis on the idea that delaying a repetition is beneficial because of changes in the *state* of the memory induced by memory consolidation and reconsolidation. Part of this changes is re-organizing a memory trace to create a more effective representation in the brain (Stickgold & Walker, 2007). Moreover, Vilberg and Davachi's (2013) finding that when an item has not consolidated (by forming greater cortical connections) the spaced repetition is not as beneficial, is consistent with this perspective.

Another difference between the reconsolidation account and previous consolidation accounts is the timescale of consolidation. The earlier consolidation accounts of the spacing effect suggested that consolidation influenced the memory on very short timescales as used in the early spacing experiments (e.g., 4.5 seconds massed vs 18 seconds spaced (Bjork & Allen, 1970)). Part of this assumption probably stems from relying on early studies that used electroconvulsive shocks (ECS) to disturb the

consolidation process (Landauer, 1969). These studies showed ECS disturbed memory on short timescales. However, some researchers have argued that ECS disturbs retrieval instead of consolidation due to the memory being recoverable with certain interventions (Miller & Marlin, 2014; Nielson, 1968). Additionally, most of the recent work on consolidation and reconsolidation uses different techniques to disturb consolidation such as using protein synthesis inhibitors or disrupting gene expression (e.g., Bekinschtein et al., 2007; Dębiec & Ledoux, 2004; Schafe & LeDoux, 2000; Suzuki et al., 2004). These studies find that a disturbance of consolidation and reconsolidation has no effect when tested seconds or hours after consolidation has been disturbed but does impact it the following day. Based on this research I believe that while a reconsolidation account is a compelling hypothesis for explaining the spacing effect over long timescales it cannot explain the spacing effect over short timescales. This viewpoint is also consistent with the behavioural spacing literature where the studies posing problems for consolidation and reconsolidation accounts of the spacing effect all use very short timescales (Bjork & Allen, 1970; Crowder, 1976).

2.3 Conclusion

By focusing on the spacing effect over short and long timescales separately, considering learning and retention and potential differences between adults and children, I have highlighted some patterns in the literature not observed in past reviews. In children, spacing enhances word and grammar learning whereas in adults it does not. Similarly, in adults spacing enhances the learning or acquisition of skills but does not enhance the learning of words or grammar. However, in both adults and children spacing generally enhances the generalization of learning and the retention of words, grammar, and skills. Accounts of the spacing effect that involve contextual variability and retrieval difficulty have some difficulty in accounting for these findings but they can be accounted for by considering participants' degree of prior experience and how that might interact with consolidation processes.

I have proposed a reconsolidation account of the spacing effect and by examining the neuroscientific evidence related to the spacing effect, we have observed some initial evidence that supports it. The initial evidence suggests spaced repetitions engage different neurophysiological mechanisms than massed repetitions. Spaced repetitions enhance the consolidation of memories to a greater extent than massed repetitions and providing time for memories to consolidate enhances the consolidation/reconsolidation of additional learning that can be fit into the same framework, resulting in faster learning and better retention. Some kind of account of the spacing effect involving consolidation and reconsolidation seems the best way to make sense of this data. Finally, there are aspects of the behavioural data that are better accounted for by a reconsolidation account than other accounts of the spacing effect such as the observation that retrieval during a spaced repetition often is easier than during massed repetitions. Therefore I believe it is worthwhile to test some of the predictions of the reconsolidation account of the spacing effect.

3 Chapter 3: Testing the influence of the strength of training on reconsolidation and the spacing effect.

3.1 Experiment 1

One way of testing the reconsolidation account is to investigate whether factors that influence reconsolidation also influence the spacing effect. In particular, previous research has indicated that increasing the strength of training generally makes memory reconsolidation more difficult to induce (Suzuki et al., 2004; Wang et al., 2009)[3]. Thus, if the spacing effect depends on memory reconsolidation, the strength of training should reduce the effect size of the spacing effect. The aim of Experiment 1 was to establish a paradigm that would allow us to test this hypothesis.

There are a number of methods that have been used to influence and observe reconsolidation in humans and I evaluated each method based on three different criteria. First, was there evidence the method tapped reconsolidation processes? Second, could the methodology also be used to demonstrate a spacing effect? Third, is it possible that a mechanism(s) other than reconsolidation explains the results? Many studies of memory reconsolidation did not meet these criteria. For example, the results of some studies could be attributed to factors other than memory reconsolidation (Gershman, Schapiro, Hupbach, & Norman, 2013; Hupbach, Gomez, Hardt, & Nadel, 2007; Sederberg, Gershman, Polyn, & Norman, 2011) while other experiments were unsuitable for showing a spacing effect (Forcato et al., 2007; Schwabe & Wolf, 2009).

[3] The strength of training is usually manipulated by increasing the number of training trials or the intensity of the training stimulus e.g. a stronger vs weaker shock.

The method employed by Strange, Kroes, Fan, and Dolan (2010) met the above criteria and, given the study provides the basis for Experiment 1, it is worth describing their experiments in detail. The starting point for Strange et al. (2010) was the observation that emotion influences memory. It is well know that stimuli associated with positive or negative emotions are better remembered than less emotionally engaging stimuli (Cahill & McGaugh, 1998; Hamann, Ely, Grafton, & Kilts, 1999). Moreover, while memory for emotional stimuli is better, memory for stimuli that precede or occur concurrently with emotional stimuli is generally impaired (Payne, Chambers, & Kensinger, 2012; Strange, Hurlemann, & Dolan, 2003). In their first two experiments Strange et al. (2010) established that emotional faces portraying fear could disrupt the consolidation process. In these experiments, participants incidentally learned 240 stem-unique nouns by categorizing them as living or non-living. Periodically, a face with a fearful expression or a neutral expression would be shown alongside the noun on the computer screen. Following this initial learning phase, participants were tested using cued recall either after a 24-hour delay (Experiment 1) or immediately (Experiment 2). The cue for recall consisted of the first three letters of the noun.

Strange et al. (2010) found that when tested after a 24-hour delay, memory for nouns immediately preceding an emotional face (e-1) was impaired relative to nouns that were preceded by a neutral face (n-1), nouns that were presented after a neutral face (n+1), and nouns that were presented after an emotional face (e+1) (Figure 7A). When tested immediately, however, emotional faces had no effect on memory recall (Figure 7B). This observation is important because if memory performance was impacted on the immediate test it would suggest that the emotional faces were interfering with retrieval or encoding rather than consolidation. Moreover, the pattern Strange et al. (2010) observed matches animal research on consolidation and reconsolidation, where substances such as protein synthesis inhibitors impair memory performance after a 24-hour delay but not when tested immediately.

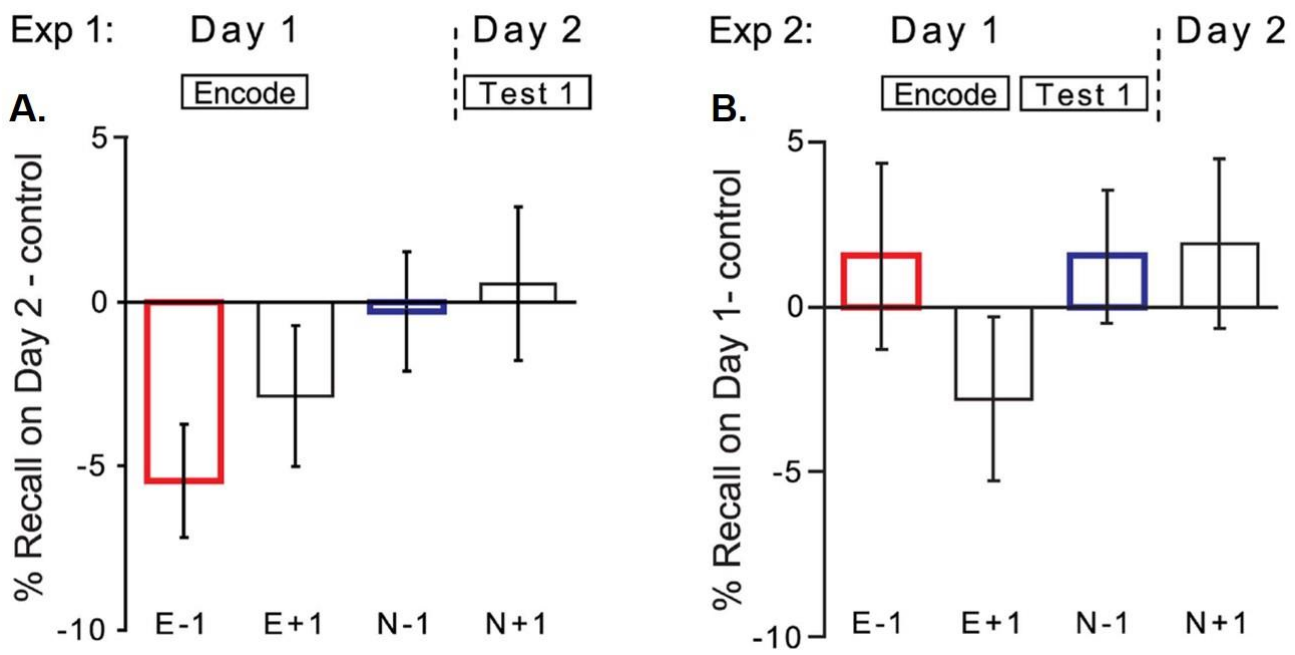

Figure 7. The methods and results for experiment 1 and 2 of Strange et al.'s (2010) consolidation and reconsolidation experiments. **A.** Recall tested after 24 hours. **B.** Recall tested immediately.

In Experiments 3 to 5, Strange et al. (2010) extended the emotional disruption effect to reconsolidation. These experiments were conducted over three days. On day 1 participants classified a series of nouns as living or non-living. On day 2 participants then completed a cued-recall test in which a fearful or neutral expression was presented every 3 to 5 trials. In Experiments 3 and 5, the face was presented alone while in Experiment 5 emotional faces were presented alongside an aversive word (e.g., murder) and neutral faces were accompanied by a neutral word (e.g., domino). Following day 2, participants in Experiment 4 completed an immediate test while participants in Experiments 3 and 5 completed the test after a 24-hour delay. Strange et al. (2010) reported that emotional faces impaired memory only for nouns that preceded them (e-1). Further, this effect occurred when the test was conducted after a 24-hour delay (Experiment 3 and 5) but not when it was conducted immediately (Experiment 4). The effect was also larger when the face was accompanied by an aversive word (d = 0.8 vs d = 0.52).

The impairment observed by Strange et al. (2010) likely occurs because, when a noun is recalled, the reconsolidation process is engaged and the memory is destabilised. Normally, it would be re-consolidated and re-stabilised, however, if an emotional face occurs after the noun has been recalled this disrupts the reconsolidation process and redirects energy towards consolidating the emotional stimulus. The Strange et al. (2010) study, therefore, provides a promising methodology to explore reconsolidation and the spacing effect. First, it allows the investigation of how different variables influence the reconsolidation process. Second, previous experiments have reported a spacing effect using methods similar to those employed by Strange et al. (2010) (Childers & Tomasello, 2002; Cull, 2000; Glenberg & Smith, 1981; Hintzman et al., 1975). Third, I can be relatively confident that reconsolidation is being explored in Strange et al. (2010) paradigm since it fits the criteria to determine if reconsolidation occurs. That is, the memory impairment is time-dependent in a manner consistent with reconsolidation.

The aim of the current experiment was to demonstrate both the spacing effect and reconsolidation using a single paradigm. The experiment consisted of three groups (Spaced, Massed and Original). In Session 1, the Spaced Group categorized a series of nouns and completed a cued-recall test. In Session 2, the Spaced Group completed a second cued-recall test with emotional and neutral faces. Finally, 48 hours later in Session 3 the Spaced Group's memory was tested. The Massed Group completed the consent form in Session 1, categorized nouns and completed the two cued-recall tests in Session 2, and was tested 48 hour later in Session 3. To ensure I could reproduce the basic findings of Strange et al. (2010) I also included the Original Group. The Original Group's procedure was identical to that used by Strange et al. (2010) with the exception that the final test was conducted after a 48-hour, rather than 24-hour delay. I hypothesised that the Spaced Group would show better overall memory than the Massed Group and that the Spaced Group and the Original Group would show worse memory for nouns that were recalled before the presentation of an emotional face compared to memory for control words or words preceding a neutral face.

Method

Participants

Sixty-two participants completed the experiment (40 Female, 22 Male; Age range: 18 – 30 years old). Thirty-eight participants were recruited via advertisements placed around the University of Otago and received $20 compensation for participating. Twenty-four participants were recruited via the Department of Psychology student experiment participation system and received course credits for completing the experiment.

Materials

Twenty neutral and fearful faces were selected from set A of the Karolinska directed emotional face set (Lundqvist, Flykt, & Öhman, 1998). The faces that were selected were those that best represented neutral and fearful expressions based on a validation study by Goeleven, De Raedt, Leyman, and Verschuere (2008). The faces were converted to greyscale and hair and clothing were removed, leaving a face on a black background (Figure 8). One hundred neutral stem-unique nouns were used as the learning stimuli. These nouns were selected from a word pool created by Warriner, Kuperman, and Brysbaert (2013) and contained valence ratings for each word on a scale of 1 to 9. Neutral words were chosen based on being stem unique and having a valence rating of 5 - 5.05 (participants in Warriner et al. (2013) were asked to respond 5 on the rating scale if they felt the word was completely neutral). Negative emotional words were also paired with each emotional face, these words were selected from Warriner et al. (2013) as words which had valence rating of 1.5 – 2.75. Neutral words were also paired with neutral faces again these had an average valence rating of 5 – 5.05. The experiment was presented on a standard windows desktop computer connected to an external microphone. The program for the experiment was developed and presented using E-Prime (Psychology Software Tools, 2013).

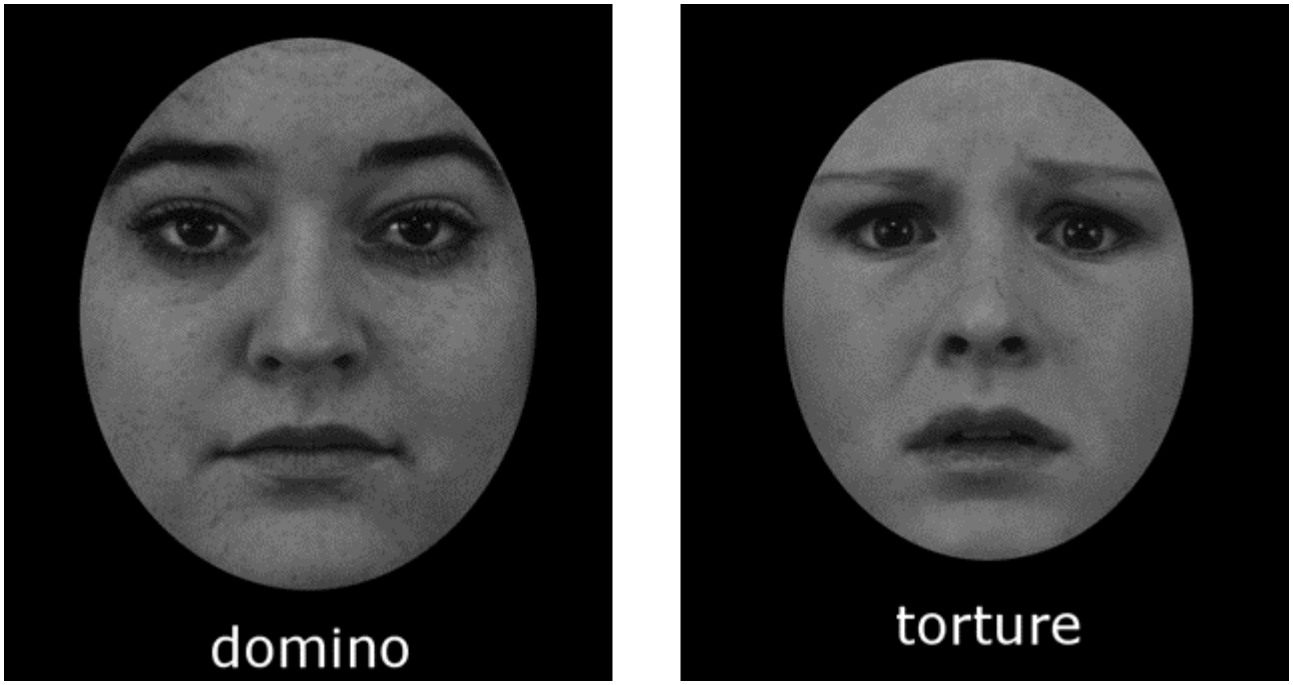

Figure 8. Examples of a neutral and emotional faces paired with a neutral and negative valenced word. These were what were used to disrupt memory reconsolidation in Experiment 1.

Procedure

Participants were randomly assigned to one of three groups; Massed, Spaced or Original. All participants completed three sessions, the first and second sessions were separated by approximately 24 hours and the second and third session were separated by approximately 48 hours. In Session 1, participants in the Spaced and Original groups categorized 100 words as living or non-living three times. Each word was presented onscreen for 1 second and this was followed by a 4-second period in which participants could make a response. In between the second and third categorization blocks participants were given a 1-minute break. After the three categorization blocks participants in the Original Group were dismissed while participants in the Spaced Group completed a cued-recall test[4]. For the cued-recall test participants were presented with the first three letters of each of the 100 words

[4] The spaced group received an extra cued-recall test because I thought it might be necessary to produce a spacing effect.

for 1 second and asked to speak the correct word into the microphone within 4 seconds. In their first session the Massed Group simply read and signed the consent form.

In Session 2 the Original and Spaced groups completed a cued-recall. After every 3 to 5 words a neutral or fearful face with a neutral or negatively valenced word was presented on the screen for 1 second. The negatively valenced word always occurred with the fearful face and the neutral word always occurred with the neutral face. After completing the cued-recall test participants in both groups were dismissed. For the Massed Group, Session 2 consisted of categorizing 100 words three times followed by two cued-recall tests. The first cued-recall test involved only recalling the categorized words while the second included the faces being presented after every 3 to 5 words.

In Session 3, the Original, Spaced, and Massed groups completed the final cued-recall test. After the cued-recall test, all groups completed a recognition test for the faces they were presented with in Session 2. The faces from Session 2 were intermixed with 22 new faces and participants were asked to respond with the *f* key if a face was familiar but they did not have the sensation of remembering it, the *r* key if they remembered seeing the face in the previous session, and the *n* key if they thought the face was new.

Data analysis

I scored the overall recall for each cued-recall test by listening to participants recordings. To explore the recall of nouns close to emotional and neutral faces and control nouns (those nouns more than 2 words away from an emotional or neutral face), I examined the memory performance of these nouns in Session 2 and 3. I initially calculated the percentage of nouns correctly recalled before or after an emotional or neutral face in Session 2 (e-1, e+1, n-1, and n+1) and the percentage of control nouns correctly recalled. I then calculated the percentage of nouns recalled in Session 2 that were correctly recalled in Session 3 for nouns before or after an emotional or neutral face and control nouns. A mixed effect Analysis of Variance (ANOVA) was used to determine the effect of the faces on noun recall.

The between-subjects factor was group (Spaced, Massed and Original). The data were also examined for normality and homogeneity using graphs, Levene's test and the Shapiro-Wilk test.

Results

Spacing Effect

Figure 9 shows the performance of all groups on the cued-recall test. There was no statistically significant difference between the Spaced and Massed groups at the first cued-recall test, $t(39) = 0.622$, $p > 0.6$, d = 0.19 reflecting the fact the procedure for this test was identical for the two groups. For the second cued-recall test the Massed Group out performed Spaced Group, $t(39) = 4.161$, $p < 0.001$, $d = 1.3$, reflecting the fact the Massed Group completed the second cued-recall test immediately while the Spaced Group completed it after a 24-hour delay. On the third and final cued-recall test the Massed Group appeared to perform better than the Spaced Group, but the difference was not significant, $t(39) = 1.811$, $p > 0.1$, d = 0.57.

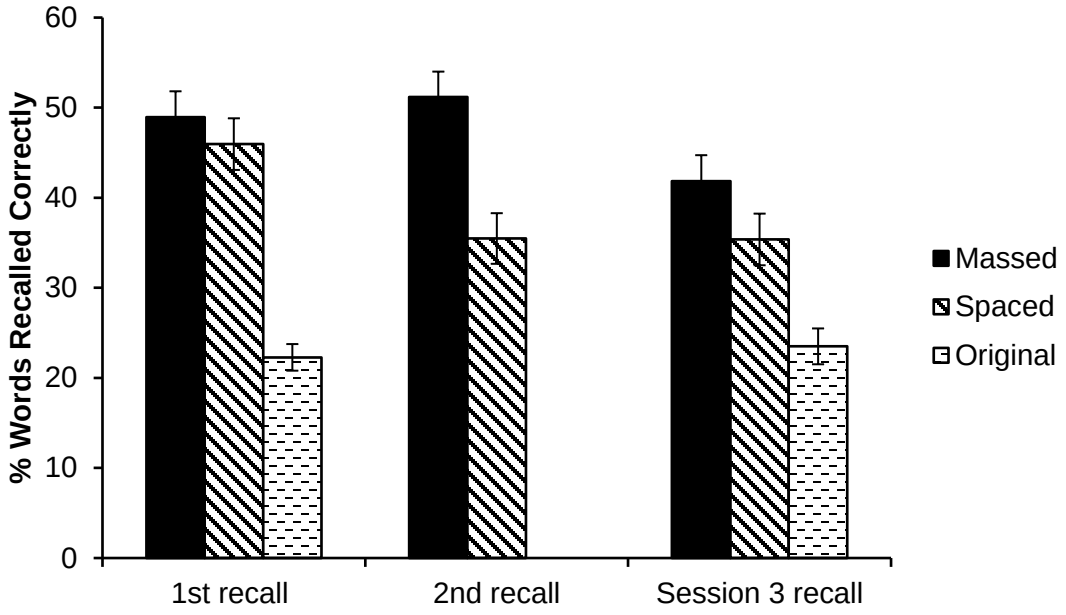

Figure 9. Mean cued-recall performance for the massed, spaced and original groups at different time points in Experiment 1. (Error bars: ± SEM)

The effect of an extra retrieval

The Original Group's procedure was identical to that of the Spaced Group, with the exception that they did not complete a cued-recall test in session 1. The extra retrieval the Spaced Group completed seems to have reduced the forgetting of the Spaced Group in comparison to the Original Group, with the Spaced Group recalling significantly more words in Session 3, $t(37) = 3.509$, $p = 0.003$, $d = 1.03$.

The effect of emotion on final recall

A mixed ANOVA was conducted to examine the influence of the emotional and neutral faces on word recall. The within-subject factor was the recall of nouns with different positions relative to an emotional or neutral face (e-1, e+1, n-1, n+1) and the between-subjects factor was design (Spaced, Massed, Original). The influence of position relative to an emotional or neutral face was not significant $F(3, 156) = 0.314$, $p > 0.8$, $\eta_p^2 = 0.06$. Nor was there a significant interaction between face and design, $F(6, 156) = 1.214$, $p > 0.2$, $\eta_p^2 = 0.046$. Overall, this indicates that the emotional faces paired with a high valence word did not seem to impact memory for the words. Figure 10 shows the means for the Spaced and Original groups for e-1, e+1, n+1 and n-1 nouns relative to control nouns. For both the Original group, a replication attempt of Strange et al. (2010), and the Spaced Group, emotional faces paired with a negative emotional word did not result in lower recall of the preceding noun compared to control nouns or nouns preceding a neutral face.

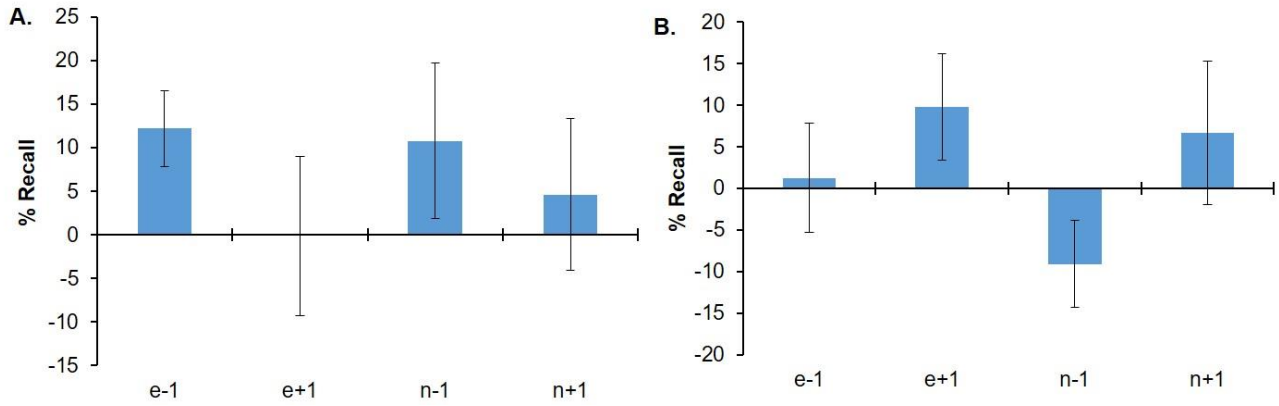

Figure 10. Experiment 1: The memory performance of different nouns was used to determine whether reconsolidation was disrupted. **A:** The Original group's memory performance in session 3 for e-1, e+1, n-1 and n+1 words relative to the recall of control nouns. **B:** The Spaced group's memory performance in for, e-1, e+1, n-1, n+1 words in session 3 relative to control nouns.

e-1: words immediately preceding an emotional face. e+1: words immediately after an emotional face. n-1: words immediately preceding a neutral face. n+1: words immediately after a neutral face. (Error bars: ± SEM).

Type of participant

In our study some of our participants were recruited via student job search and advertisements around the university and received $20 for participation whereas others were recruited from first and second year psychology classes and were given course credits for participation. To examine whether these two types of participants may have influenced my results I conducted an ANOVA with participant type (credit or cash) and design (Spaced, Massed, Original) as the independent variables and session 3 recall as the measured variable. I found a significant effect of design, $F(2, 53) = 10.117$, $p < 0.001$, $\eta_p^2 = 0.276$ and participant type, $F(1, 53) = 4.136$, $p < 0.05$, $\eta_p^2 = 0.072$, but no significant interaction, $F(2, 53) = 0.006$, $p > 0.6$, $\eta_p^2 = 0.017$. The significant effect of participant type reflected the fact that the participants who received cash had a higher mean recall on the final test than participants who received credit (M = 35.70 vs M = 29.00).

Discussion

The results of Experiment 1 were not consistent with my hypotheses. First, I failed to replicate Strange et al.'s (2010) finding that emotional faces disrupt the recall of the words preceding them.

Indeed, there were no reliable differences in recall of words preceding emotional faces, neutral faces, or control words. Second, I expected participants in the Spaced Group would show higher overall recall than participants in the Massed Group. In contrast, participants in the Massed Group showed *better* recall than participants in the Spaced Group. This result did not reach the threshold of statistical significance but this probably partly due to a small sample size.

Why did I fail to replicate Strange et al. (2010)? One possibility is that the current study represents a genuine failure to replicate. Recently, Aarts et al. (2015) set out to replicate 100 psychological studies. When the original studies were conducted, 97% obtained a statistically significant result, however, only 36% of the replication studies obtained statistically significant results. This finding suggests that when an independent researcher attempts to replicate a study there is a low baseline probability of obtaining a significant result. The results of the current study could simply be a reflection of this general observation.

Alternatively, it is possible that small differences between the current study and Strange et al. (2010) led to the null result. These differences include that I used a different set of nouns to Strange et al. (2010). The nouns, however, did conform closely to the nouns described by Strange et al. (2010) in their study because they were similar kinds of nouns, stem unique, the valence of the nouns was carefully controlled and they were similar in length. Also, rather than selecting faces randomly from the Karolinska face set, I chose the faces from the set that best reflected neutral and fearful expressions (Goeleven et al., 2008). One final difference is that, due to not having access to the instructions provided to the participants by Strange et al. (2010), my instructions may have differed in some way. Of these three differences, it seems unlikely using different nouns or a different method of selecting faces would produce such different results, as the memory and reconsolidation processes are unlikely to be affected in any significant way by these changes. Subtle differences in instructions, however, can impact these processes. A number of studies have demonstrated that how participants perceive the

significance or relevance of a task influences the consolidation processes (Abe et al., 2011; Fischer & Born, 2009; Stickgold & Walker, 2013; Sugawara, Tanaka, Okazaki, Watanabe, & Sadato, 2012; Wilhelm et al., 2011). In terms of the current experiment, it is possible that the instructions influenced how significant participants thought the faces to be and this in-turn may have influenced the reconsolidation process.

A second question is why the Massed Group recalled more words than the Spaced Group? This is likely a result of the particular combination of spacing interval, method of learning, and retention delay used in the current experiment. Although the incidental learning method employed in the current study (i.e., categorizing the nouns as living or non-living) has been used in previous studies of the spacing effect, these studies tended to use much shorter spacing intervals than those employed in the current experiment (Glenberg & Smith, 1981; Shaughnessy, 1976; Toppino & Bloom, 2002; Verkoeijen et al., 2005). Incidental methods of learning often have lower levels of recall than intentional methods of learning (Postman & Phillips, 1954; Roediger, 2008) and, in the current study, participants in the Spaced Group displayed a large drop in recall between Session 1 and Session 2. Additionally, the spacing effect is influenced by the retention interval. Generally speaking, a longer retention interval means the spacing effect is *more* likely to emerge. Most studies with spacing intervals of 24 hours or more have used retention delays longer than the 48 hours used in the current study. It is possible that, while the Massed Group showed initially better memory than the Spaced Group, a longer retention delay (e.g., 1 week) may have revealed a spacing effect.

3.2 Experiment 2

As noted above, it is possible I did not replicate Strange et al. (2010) due to differences in the instructions provided to participants. One unique aspect of the Strange et al. (2010) experiment was that they used a within-subjects design, where emotional but not neutral faces promoted a specific memory disruption within participants. In many studies on reconsolidation, however, a between-subjects design has been used and a more general disruption is generally reported (Forcato et al., 2007; Schwabe & Wolf, 2009; Walker et al., 2003). Perhaps, the emotional faces in Experiment 1 did disrupt reconsolidation but, because of the differences in instructions, it was no longer specific to the noun just preceding the faces and promoted a more general memory disruption. Based on this reasoning, in Experiment 2 I used a between-subjects design which may be more robust to minor differences in instructions and participants perception of the task.

In addition, I sought another way to disrupt consolidation. Several recent experiments by Brown and Robertson (2007) have reported that memory consolidation of a word-list is disrupted by the subsequent learning of a motor skill. Furthermore, greater learning of the motor-skill task was associated with greater impairment in memory for the word-list. Other studies have disrupted the consolidation and reconsolidation process by having participants learn a short story (Schwabe & Wolf, 2009). Together, these studies suggest that there may be many different ways of disrupting the consolidation and reconsolidation of memory. The key is that these methods involve new learning.

In Experiment 2, participants in the Spaced Group *intentionally* learned Swahili-English paired associates (Session 1), repeated the words 24-hours later (Session 2), and had their retention tested after a 1-week delay (Session 3). Participants in the Massed Group followed the same procedure with the exception that Session 1 and Session 2 were condensed into a single session. Participants in the Disruption Group were identical to the Spaced Group, with the exception that they learned a series of

face-name paired associates in Session 2. In theory, the face-name paired associates should disrupt the reconsolidation of the Swahili-English pairs because they both use resources within the declarative memory system (Brown & Robertson, 2007; Schwabe & Wolf, 2009). I hypothesised that I would observe a spacing effect, with the Spaced Group showing less forgetting than the Massed Group, and a disruption effect with the Spaced Group showing less forgetting than the Disruption Group.

Method

Participants

Sixty-four participants completed the experiment (41 females and 23 males; Age range: 18 – 49 years old). Participants were recruited via advertisements placed around the University of Otago and via the Department of Psychology student experiment participation system. Participants either received $20 compensation at the end of their third session or course credits.

Materials

Forty Swahili-English word-pairs were randomly selected from Nelson and Dunlosky (1994) to form a list of paired-associates as the primary learning materials. Face-name paired-associates were formed by selecting faces from the Karolinska face set and pairing them with an English name generated by the experimenter. Half of the faces were male and half were female. The experiment was presented on a standard desktop computer. The program for the experiment was developed and presented using E-Prime software (Psychology Software Tools, 2013).

Procedure

Participants were randomly assigned to the Spaced, Massed or Disruption group. In Session 1, participants in the Spaced and Disruption groups were shown 40 Swahili-English paired associates. Each word-pair was shown for 5 seconds. Following this, participants were presented with the Swahili word and asked to type the English equivalent (i.e., the paired associate). Participants had unlimited time to type the word. If participants translated the word correctly they were told their answer was correct and the word-pair was presented for 4 seconds. If participants translated the word incorrectly

they were told their answer was incorrect and shown the correct pairing for 4 seconds. After being tested on all 40 paired associates, participants were re-tested on the paired associates they had translated incorrectly. The session was completed after participants had responded to all paired associates correctly. Similar to Experiment 1, on the first session participants in the Massed Group simply signed the consent form and were dismissed.

In Session 2, participants in the Spaced and Disruption groups completed two cued-recall tests on all 40 paired associates. Again, participants had unlimited time to respond and feedback was provided. After the two cued-recall tests, participants in the Spaced Group completed 120 simple addition and subtraction equations while participants in the Disruption Group were presented with the 30 face-name pairs. Following this, participants in the Disruption Group completed a two cued-recall tests where they were presented with the face and had to recall the name. The testing procedure was identical to that used for word pair-associates in Session 1. In Session 2, the Massed Group completed the tasks completed by the Spaced Group in Sessions 1 and 2.

Session 3 was conducted one week after Session 2 for all participants. The Spaced and Massed groups were tested on the Swahili-English word-pairs and the Disruption Group was tested on the Swahili-English word pairs and then the face-name pairs.

Results

Outliers

One subject was excluded from the Massed Group because, after the initial encoding, they scored 80% correct when tested on the first 40 items. The mean across all groups at the same time point was 14.01% and the outlier participants score was 5.5 standard deviations from the mean.

Performance prior to the final memory test

I examined whether there were any differences between the three groups prior to the final retention test. By examining the first 40 trials of the cued-recall test conducted during the learning

phase, I was able to determine the percentage of word-pairs participants could remember after viewing each pair for 5 seconds. There was a trend toward the Spaced Group remembering fewer word-pairs than the Massed and Disruption groups, $F(2, 60) = 2.801$, $p = 0.069$, $\eta_p^2 = 0.085$ (Figure 11). I also examined how many trials each group required to correctly recall each word-pair, there was no difference between groups, $F(2, 60) = 2.004$, $p = .144$, $\eta_p^2 = 0.063$.

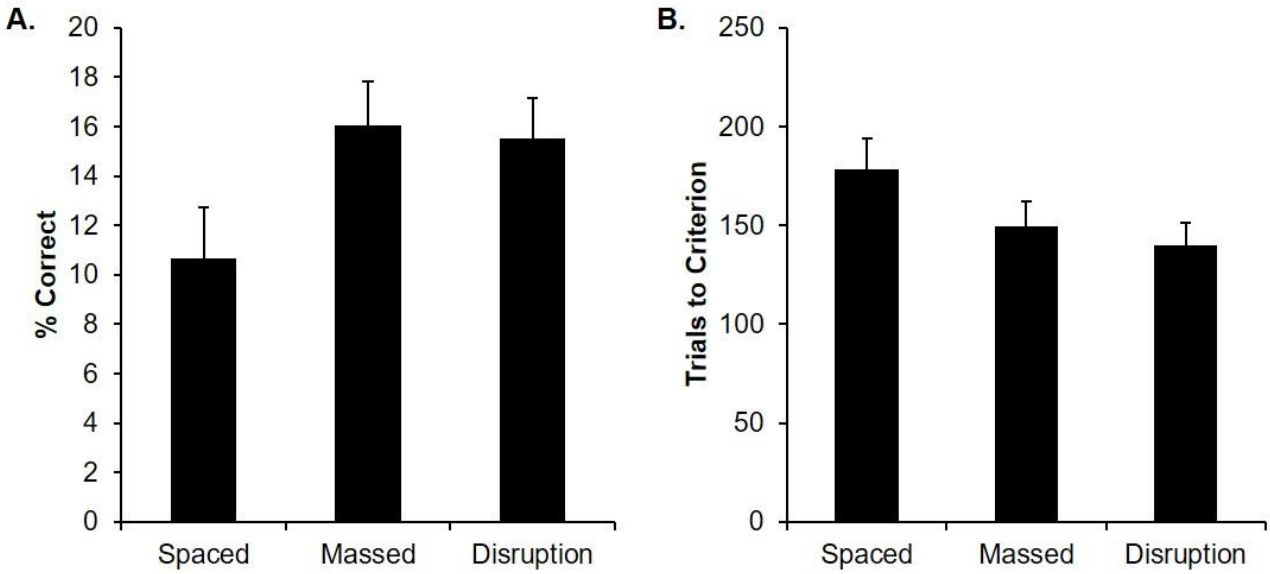

Figure 11. Participants learning performance in Session 1 of Experiment 2: **A**. Mean percentage correct after one presentation of each word. **B.** Mean number of trials to criterion for the Swahili-English word-pairs. (Error bars: ± SEM).

Cued-recall tests 1 and 2

For the first cued-recall test, an ANOVA revealed a significant effect of group, $F(2, 60) = 4.782$, $p < 0.02$, $\eta_p^2 = 0.137$. Tukey's post-hoc tests revealed that there was a significant difference between the Spaced Group and the Massed Group, but not the Massed Group and the Disruption Group. On the second cued-recall test, although the recall of the Spaced Group appeared lower than the Massed Group and Disruption Group, there was no main effect of group, $F(2, 60) = 1.847$, $p > 0.1$, $\eta_p^2 = 0.051$ (Figure 12).

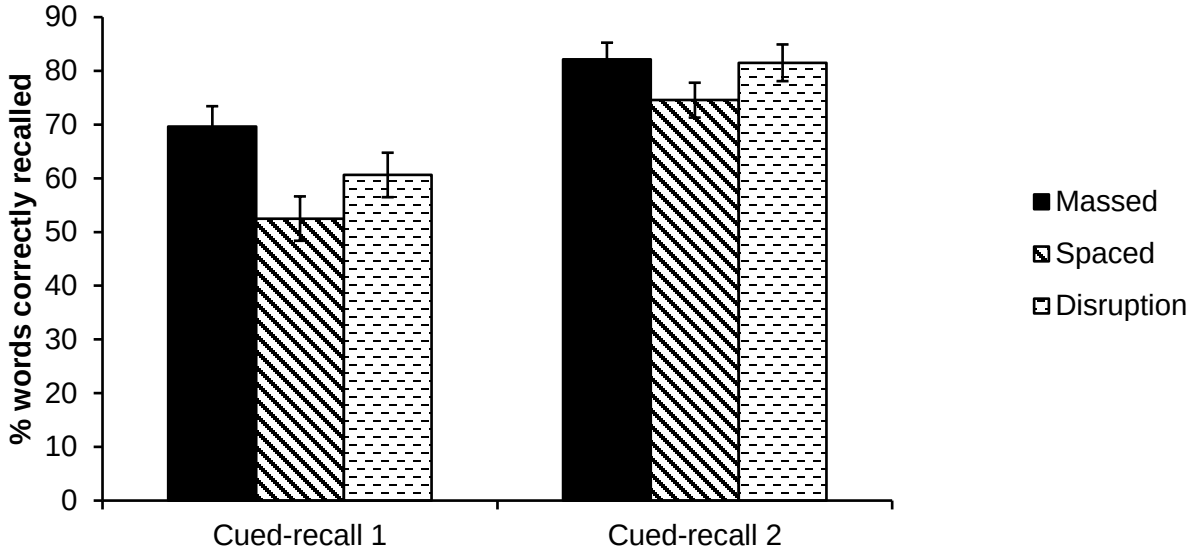

Figure 12. Percentage of word-pairs correctly recalled in the two cued-recall tests after learning for the massed, spaced and disruption groups (Session 2 of Experiment 2). (Error bars: ± SEM).

Mean recall after a 1 week delay

There was a significant difference between the groups for the percentage of words correctly recalled on the retention test, $F(2, 60) = 2.577$, $p < 0.05$, $\eta_p^2 = 0.107$. In terms of mean performance, the Disruption Group displayed a slightly higher recall than the Spaced Group ($p > 0.05$) and the Spaced Group displayed slightly higher recall than the Massed Group (Figure 13) but none of these comparisons were statistically significant all *p's* > 0.05. Since learning the face-name pairs did not disturb memory performance in the Disruption Group, I collapsed across the Spaced and Disruption groups. When combined, there was a significant spacing effect on mean percentage recall on the retention test, $F(1, 60) = 4.997$, $p < 0.05$, $g = 0.598$.

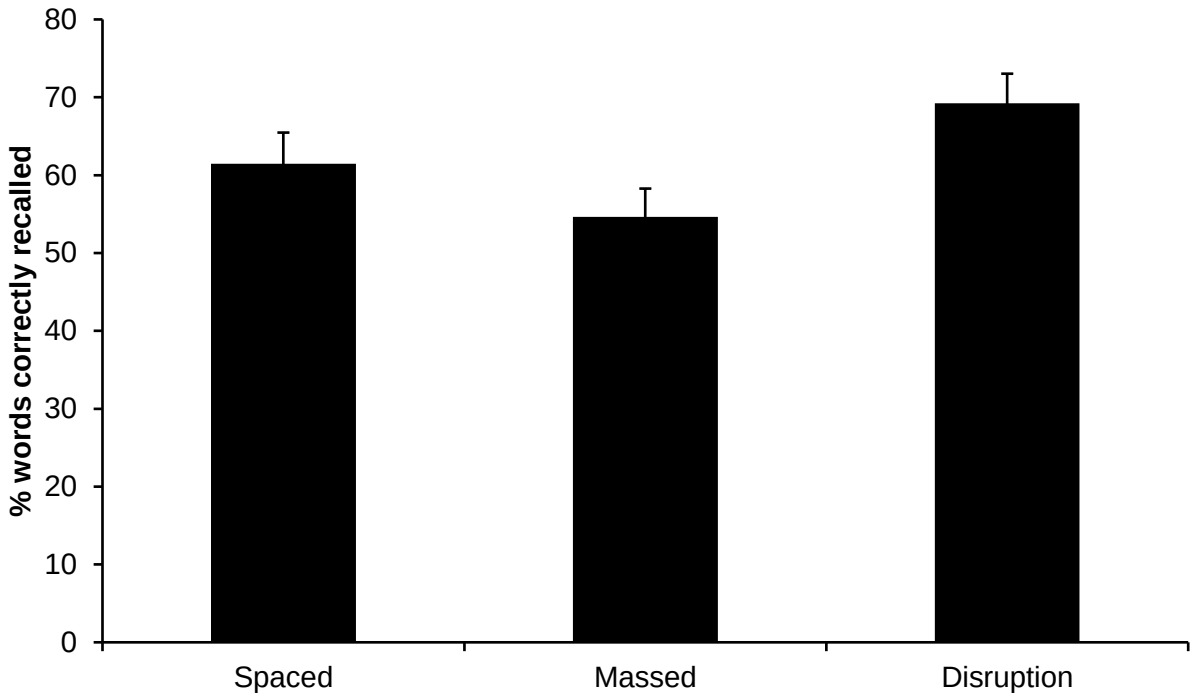

Figure 13. Percentage of word-pairs correctly recalled in the retention test completed in Session 3 of Experiment 2. (Error bars: ± SEM)

Amount of forgetting

One reason I did not obtain a significant group effect in our initial ANOVA (i.e., when the data were analysed with the Spaced Group and the Disruption Group separated) was due to the large variation in memory performance within each group. One way to reduce the within group variation is to obtain a measure of forgetting by subtracting participants performance on the retention test from their performance on the second cued-recall test. I conducted an ANOVA with the amount forgotten as the dependent variable and found a significant effect of group, $F(2, 60) = 13.406$, $p < 0.001$, $\eta_p^2 = 0.308$ (Figure 14). I then conducted Tukey's post-hoc tests and found a significant difference between the Spaced Group and Massed Group ($p < 0.001$, $d = 1.33$) and the Disruption Group and the Massed Group ($p < 0.001$, $d = 1.42$) but no significant difference between the Spaced Group and Disruption Group ($p > 0.9$, $d = 0.078$).

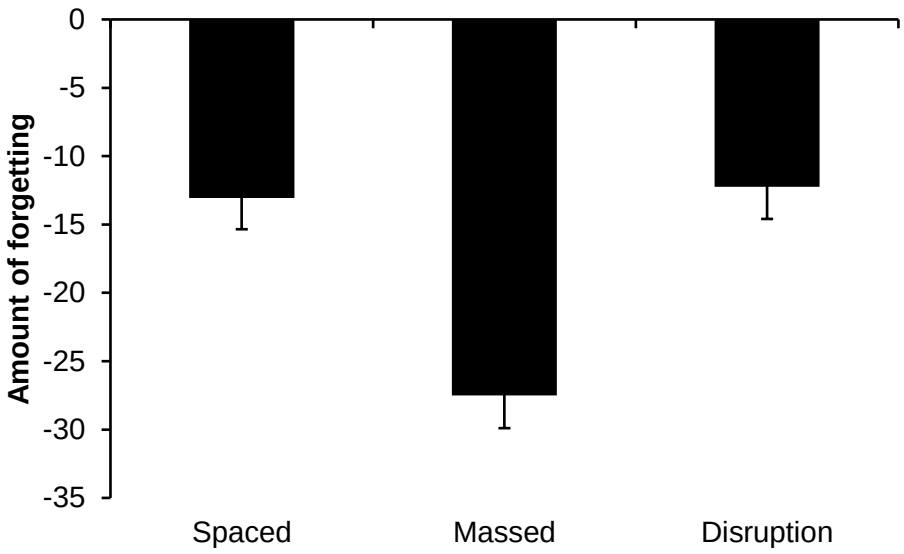

Figure 14. Amount of forgetting for the spaced, massed and disruption groups in Experiment 2. Amount of forgetting is calculated as performance on the final test (Session 3) – performance on cued recall test 2 (Session 2). (Error bars: ± SEM).

The Disruption Group's performance on face-name pairs

There was no significant difference in the initial recall of face-name (M = 18.62, *SD* = 9.05) and Swahili-English (*M* = 15.50, *SD* = 7.33) word pairs, $t(1,19) = 1.210$, $p = 0.241$. I was also interested in whether performance on the retention test differed between Swahili-English pairs and face-name pairs because the spacing of repetitions differed for these types of stimuli. A greater percentage of Swahili-English word-pairs were correctly recalled on the retention test (*M* = 69.25, *SD* = 16.92 vs. *M* = 36.37, *SD* = 16.66), $t(1, 19) = -7.58$, $p < 0.001$, $d = 1.95$, potentially reflecting a within-subjects spacing effect.

Discussion

In Experiment 2, I hypothesised that the Spaced Group would show less forgetting than the Massed Group and the Disruption Group. The Spaced Group did show less forgetting than the Massed Group, illustrating a spacing effect. However, the Spaced Group showed a similar amount of forgetting to the Disruption Group, indicating that learning face-name pairs did not disrupt the reconsolidation of the Swahili-English word pairs. What could explain the lack of memory disruption? One possibility is

that the type of test influences whether a memory disturbance is observed. The studies referenced above demonstrate that learning quite different materials can disrupt memory reconsolidation when free-recall is used during testing. In the current study, I used cued-recall and found that similar declarative memory materials did not disrupt memory. Additionally, consistent with this perspective, other reconsolidation studies that use test methods other than free-recall (e.g., cued-recall) use materials very similar to the initial learning to create the memory disturbance (Forcato et al., 2007; Soeter & Kindt, 2011; Walker et al., 2003; Wichert, Wolf, & Schwabe, 2011). This observation suggests that the impairment in declarative memory caused by a procedural task observed by Brown and Robertson (2007) might not be due to their being an overlap in the consolidation of declarative and procedural memories.

A more probable explanation for the Brown and Robertson (2007) results is that there was contextual interference produced between the free-recall and motor-skill tasks. Generally, free-recall tasks use English words that participants are already familiar with. Thus, when participant are in the learning stage of a free-recall task participants are retrieving words from their memory and binding them to that particular context and to other words in the list. To later recall the words participants use cues from the context to know which words they should be retrieving. The free-recall task is, therefore, heavily dependent on the context. It may be the case that in Brown and Robertson's experiment, because the motor-skill task was learned in the same physical context and a similar temporal context (i.e. the motor skill task was learned immediately after learning the word list), that when the participants are using context cues to recall the list of words they spontaneously recall elements from the sequence and this interferes with their recall of the word list.

A similar explanation can be applied to the Schwabe and Wolf (2009) study. Participants recalled an episode of their life, which probably led to new contextual cues being integrated into the memory. Some participants then learned a story within the same context, whereas others completed a distractor task. Later, when asked to recall the episode of their life, participants probably used contextual cues

that they had integrated into the memory the previous day. Thus, again, the overlap between the contextual cues used for the episode and those used for the story may have produced interference. One finding that may argue against this explanation is that the memory impairment was observed after a 24-hour delay but not on the immediate test. One explanation for this is that when the memory was tested immediately the memories were relatively strong and easily retrievable and thus could overcome any interference. Cued-recall tasks are less dependent on the context than free-recall tasks and, therefore, might not be susceptible to contextual interference. The current experiment supports this conclusion as learning face-name paired associates immediately after the repetition of Swahili-English paired-associates had no effect on memory recall.

3.3 Experiment 3a

In the previous experiment I found that two different types of paired-associates did not produce a disturbance of memory reconsolidation. I suggested that this may be due to consolidation and reconsolidation being highly specific and, as a result, only similar types of materials produce a disturbance of these processes. This is consistent with the fact that past studies have used very similar materials to disturb the reconsolidation process (Forcato et al., 2007; Walker et al., 2003; Wichert et al., 2011). Using more similar materials, however, introduces the possibility that any impairment observed is due to interference rather than disturbing memory reconsolidation.

Many studies of interference have used the A-B A-C paradigm. In this paradigm participants first learn a list of paired-associates. This first list is generally referred to as "A-B" with A representing the stimulus term and B representing the response term. If participants are in the interference condition they then learn what is referred to as the A-C list. In this list the stimulus term of each paired associate is the same as the A-B list but the response term is a new word. The interference is therefore caused by the stimulus terms of lists A-B and A-C overlapping and being associated with different responses. In the A-B A-C paradigm the interference condition is compared with a control condition in which participants learn two different lists with no overlapping terms (i.e., A-B and C-D).

In Experiment 3a I aimed to determine if more similar materials would disturb the consolidation of Swahili-English paired associates without producing interference. I used an A-B C-D design. There were three groups. Participants in the List-1-Only Group learned a list of Swahili-English paired-associates and then came back the next day for a recall test. Participants in the Disruption Group learned a list of paired-associates and, after a brief break, learned a second list of paired-associates with different stimulus and response terms. Following this, participants had their recall tested the next day. Finally, participants in the Immediate-Test Group learned a paired-associate list, learned a second list, and then had their memory for the first list tested immediately. I hypothesised that recall for the List-

1-Only Group would be higher than that for the Disruption Group but no different from the Immediate-Test Group.

Method

Participants

Sixty participants completed this experiment (39 Female, 21 Male; Age range: 18 – 37 years old). Participants were recruited via advertisements placed around the University of Otago and via the Department of Psychology student experiment participation system. Participants either received $20 compensation at the end of their third session or course credits.

Materials

Two lists of Swahili-English paired-associates were constructed, some word-pairs were selected from Nelson and Dunlosky (1994) and other word-pairs were generated by the experimenter selecting English words related to animals or food and finding the Swahili translation. Each list had 15 word-pairs. In one list all of the word-pairs were related to food and in the other they were all related to animals. A theme was chosen for each list so that it would be easier for participants to distinguish the word lists. A standard windows desktop computer was used to present the experiment, while the experiment was developed and presented using E-Prime software (Psychology Software Tools, 2013).

Procedure

Participants assigned to the List-1-Only Group were randomly assigned to either the animal or the food list as list 1. For each list, each pair was presented for 5 seconds. After all 15 word pairs were presented; participants began the process of learning each Swahili word-pair to a criterion of one correct recall. Participants were presented with the Swahili word as the cue and were asked to type the English translation. The Swahili cue remained on screen until the participant typed a response and pressed enter. After entering a response participants were told whether they were correct or incorrect

and presented with the correct Swahili-English word pair for 4 seconds. After the participants had typed a response for all word-pairs, the correct word-pairs were eliminated from the list and the incorrect word-pairs were presented again in the same manner. After responding to all word pairs correctly, participants were tested on all 15 word-pairs (Test 1). No feedback was provided during the test. Participants returned 24-hours later and completed another cued-recall test on the word pairs (Test 2).

Participants in the Disruption and Immediate-Test group were randomly assigned either the animal or food list, learned it to a criterion of one correct recall for each word-pair, and were then tested on the list they had learnt. After list 1 was tested, the Disruption and Immediate-Test groups were presented with a new set of instructions. These instructions said that they would now learn a second list that would be tested in tomorrow's session. They were then presented with the other list, learned it to a criterion of one correct recall, and were then tested on the list they learnt. Following this, the Disruption Group was dismissed while the Immediate-Test Group completed a second test on the first list they learnt. Both groups were then tested the next day on list 1 and list 2.

Results

Learning performance

Firstly, I conducted a one-way ANOVA to examine the effect of group on recall of list 1 after initial encoding. There was no significant effect of group on the percentage of word-pairs correctly recalled after initial encoding, $F(2, 57) = 0.637$, $p = 0.533$, $\eta_p^2 = 0.022$. Additionally, in order to determine whether recall after encoding differed for list 1 and list 2 I conducted a mixed model ANOVA with List (1 and 2) as within-subjects factors and group (Disruption, Immediate-Test) as the between subjects factor. There was a significant main effect of List, $F(1, 37) = 5.729$, $p < 0.03$, $\eta_p^2 = 0.134$, but no main effect of Group or List by Group interaction both *p's* > 0.05. This indicates that after a single presentation participants remembered more of List 2 than List 1 (Figure 15). I also conducted a between-subjects ANOVA with the number of trials to criterion as the measured variable

and group (List-1-Only, Disruption and Immediate-Test), there was no significant difference between groups, $F(2, 57) = 0.575$, $p > 0.5$, $\eta_p^2 = 0.20$, for the number of trials required to learn list 1. For the Disruption and Immediate-Test groups there was again a significant effect of List, $F(2, 37) = 11.695$, $p < 0.005$, $\eta_p^2 = 0.240$, indicating that List 2 was learned faster than List 1.

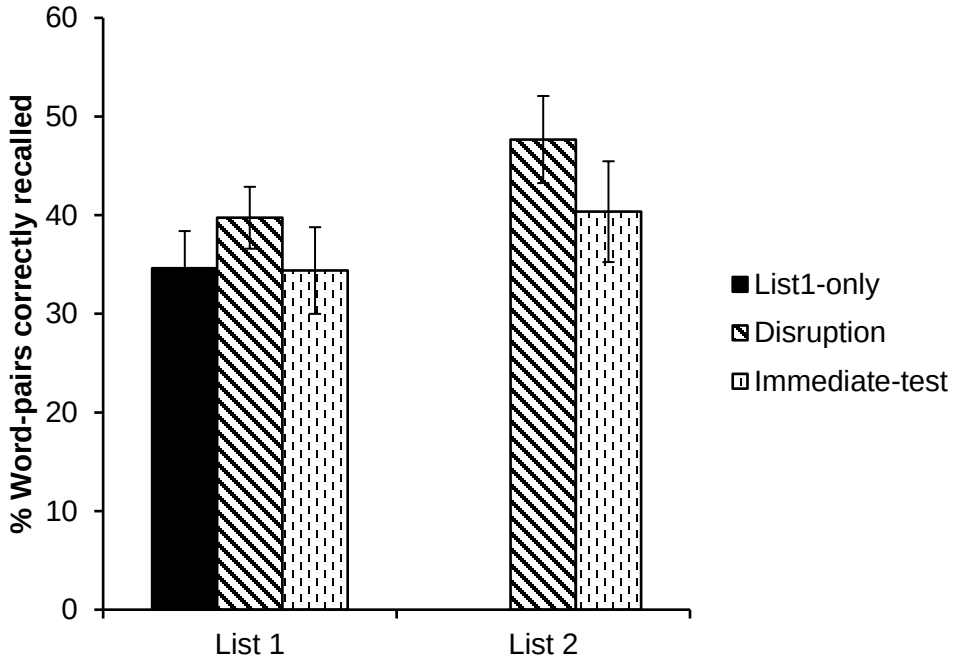

Figure 15. Percentage of word-pairs correctly recalled after initial encoding (Session 1, Experiment 3a). (Error bars: ± SEM).

Test 1, 2 and the amount forgotten

There was no significant difference in participants performance on Test 1, $F(2, 57) = 0.283$, $p > 0.7$, $\eta_p^2 = 0.010$ or on Test 2, $F(2, 57) = 0.674$, $p > 0.5$, $\eta_p^2 = 0.023$. A more refined analysis is to examine the amount forgotten. This was calculated by taking an individual's Test 1 score and subtracting it from their Test 2 score (Session 2 score – Session 1 score, Figure 16). An ANOVA with List 1 forgetting as the dependent variable revealed a significant main effect of Group, $F(2, 57) = 3.159$,

$p < 0.05$, $\eta_p^2 = 0.100$. Tukey's post-hoc tests revealed that the List-1-Only Group forgot significantly less than the Disruption Group ($p < 0.05$, $d = 0.85$). Additionally, the Immediate-Test Group was not significantly different to the List-1-Only group ($p > 0.3$, d = 0.34) or the Disruption Group ($p > 0.3$, d = 0.41).

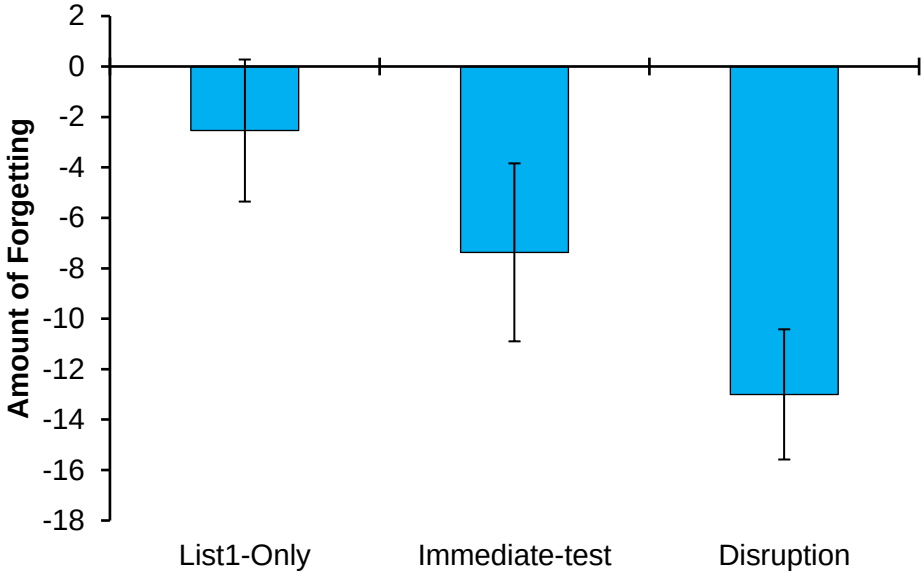

Figure 16. Mean amount of forgetting (Session 2 score – Session 1 score) for the List1-Only, Immediate-Test and Disruption group in Experiment 3a. (Error bars: ± SEM).

Theme of list

List 1 and 2 could be either an animal or a food list. I examined whether participants performed differently on these lists. There was no significant differences in the number of trials to learn the animal ($M = 34.91$, $SD = 13.73$) or food lists ($M = 36.93$, $SD = 10.29$), $t(58) = 0.638$, $p > 0.5$, $d = 0.08$. Nor was there any significant difference between the animal ($M = 34.76$, $SD = 16.27$) and food list ($M = 37.54$, $SD = 17.47$) in the initial percentage of correct recall after encoding, $t(58) = 0.635$, $p > 0.5$, $d = 0.16$. Participants who received the food list ($M = 80.63$, $SD = 15.09$) did, however, recall a greater percentage of words correct than participants who received the animal list ($M = 71.90$, $SD = 18.18$) on Test 1, $t(58) = 2.030$, $p < 0.05$, $d = 0.52$. Participants who received the food list ($M = 76.25$, $SD = 16.32$) also showed significantly better recall on Test 2 than participants who received the animal list

(M = 60.71, SD = 19.23), $t(58) = 3.385$, $p < 0.005$, $d = 0.87$. Finally, participants who received the animal list also showed a greater amount of forgetting than the participants who received the food list and this was just above the threshold of significance, $t(58) = 1.985$, $p = 0.056$, $d = 0.498$ (Figure 17).

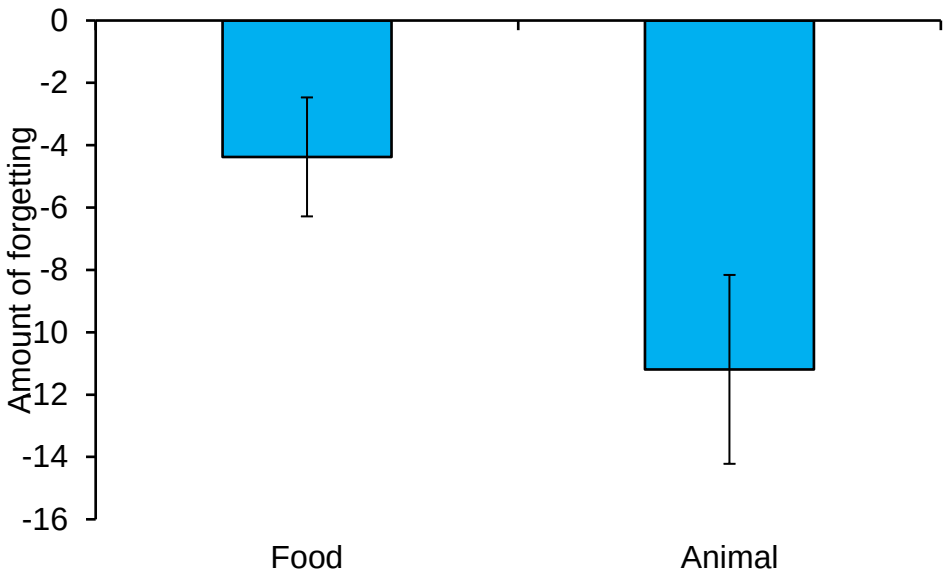

Figure 17. Mean amount of forgetting for the food and animal lists (Experiment 3a). (Error bars: ± SEM).

Discussion

The aim of the current experiment was to determine whether the consolidation of Swahili-English word-pairs would be disturbed by the learning of a second set of word-pairs. The ideal criterion for the disturbance of consolidation, based on other neurophysiological and behavioural studies, was for the disturbance to have no immediate effect on memory but have an effect when the memory is tested after a 24-hour delay. The results of the current study partially fulfilled this criterion. The Disruption Group was significantly different from the List-1-Only Group and, the Immediate-Test Group was not significantly different from the List-1-Only Group. The current study does not perfectly fulfil the

criterion because the Immediate-Test Group's performance was intermediate and not significantly different from the Disruption Group.

There are three plausible explanations for our results. Synaptic consolidation involves changes in the morphology of synapses and synaptic potentials and is thought to occur on a timescale of minutes to hours, whereas systems consolidation occurs primarily during sleep and is generally associated with memory replay and the shift of memories from the hippocampus to the cortex (Frankland & Bontempi, 2005; Mednick, Cai, Shuman, Anagnostaras, & Wixted, 2011). One explanation of our findings is that learning the additional set of Swahili-English word-pairs caused a small initial disturbance of synaptic consolidation and, overtime, as the second list continued to disturb synaptic and systems consolidation of the first list there was additional forgetting. A second possibility is that learning the second list produces a small amount of interference. On the immediate test, the influence of the interference creates a small non-significant deficit and on the delayed test the combination of a small amount of interference and disruption of consolidation produced the larger deficit I observed. A third possibility is that the entire memory deficit produced on the delayed test is produced by interference but when an immediate test is conducted the initial memory is stronger so the interference is less effective in disrupting the memory. Interference with retrieval seems unlikely because the words of the two lists were semantically quite different and there was little auditory similarity when comparing the sounds of the Swahili words in the two lists. Moreover, past research has reported that a memory deficit produced by retroactive interference decreases or remains stable over time (Ceraso & Henderson, 1965; Drosopoulos et al., 2007; Postman, Stark, & Fraser, 1968) whereas in our study the memory deficit increased over time.

I also examined whether there was a difference in the forgetting of the animal or food lists. Participants forgot significantly more words from the animal list than the food list and this did not interact with group membership. Interestingly, the advantage of the food list over the animal list seems to be relatively specific to memory performance after a delay period. There was no significant

difference in the number of trials to learn the animal and food lists or in the percentage correct after the initial encoding. There was, however, a significant difference on Tests 1 and 2. Test 1 for both groups occurred after learning to criterion so, for the particular words being tested, this constitutes a small delay and Test 2 generally occurs after 24 hours. Based on these results and an examination of the effect sizes at different time points (Figure 18) one can see that over time performance on the two lists diverges.

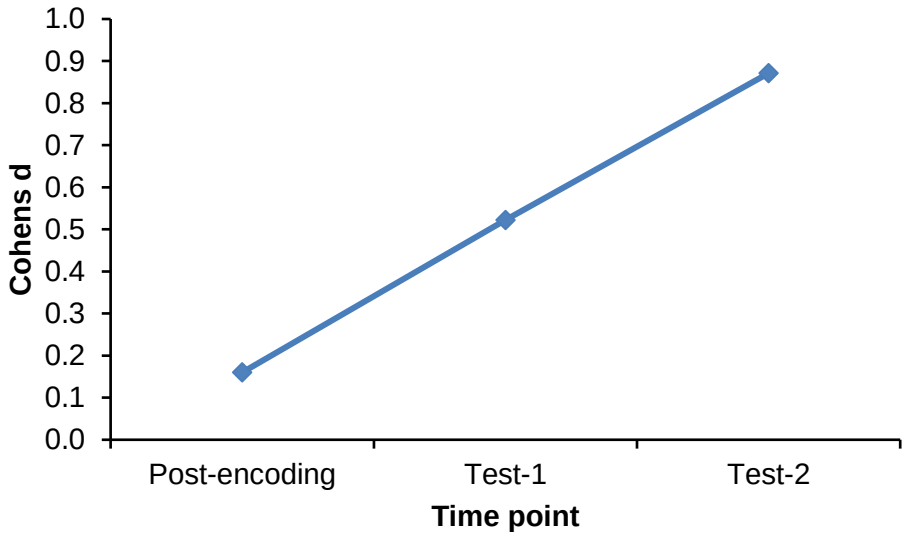

Figure 18. Cohen's d effect size for the difference between recall of the food and animal list at different time points in Experiment 3a. Participants always show superior performance on the food list relative to the animal list.

3.4 Experiment 3b

Experiment 3a suggested that learning a second list of Swahili-English paired associates disturbed the consolidation of the first list. The effect, however, may also be due to interference. The aim of Experiment 3b was to determine whether interference or consolidation was producing the memory disturbance observed in Experiment 3a. Part of the core idea of the consolidation process is that the memory is in an initially unstable state where it is susceptible to interference and disruption but, over time, it consolidates and becomes less susceptible to interference and other disturbances. In Experiment 3b, instead of participants learning a second list immediately as they did in Experiment 3a, they learned the second list on day 2. I hypothesised that the recall for participants who learned the second list on day 2 would not be significantly different from participants who learned only one list. That is, because the second list is learned after the first list would have had sufficient time to consolidate, it should not be as susceptible to interference or disturbances of memory consolidation.

Method

Participants

Nineteen participants completed the experiment (Male = 4, Female = 15; Age range: 18 – 37 years old). Participants were recruited via advertisements placed around the University of Otago and via the Department of Psychology student experiment participation system. Participants either received $20 compensation at the end of their third session or course credits.

Materials

All of the materials were the same as in Experiment 3a.

Procedure

In Session 1, participants learned a list to a criterion of 1 correct recall for each word and were dismissed. In Session 2, participants learned list 2 and completed a cued-recall test. Participants were then tested on list 1 and completed another test of list 2.

Results

Learning performance

I combined the data of Experiment 3a and 3b to allow a direct comparison between the different conditions. Given participants in Experiment 3b completed the experiment at a different point in time I tested whether they displayed similar performance on List 1 to participants in Experiment 3a. There was no significant difference in the percentage of word-pairs correctly recalled, $t(77) = 2.788$, $p = 0.064$, $d = 0.483$ (Figure 19A). Participants in Experiment 3b did, however, require a greater number of trials to reach criterion, $t(77) = 3.385$, $p < 0.05$, $d = 0.63$ (Figure 19B).

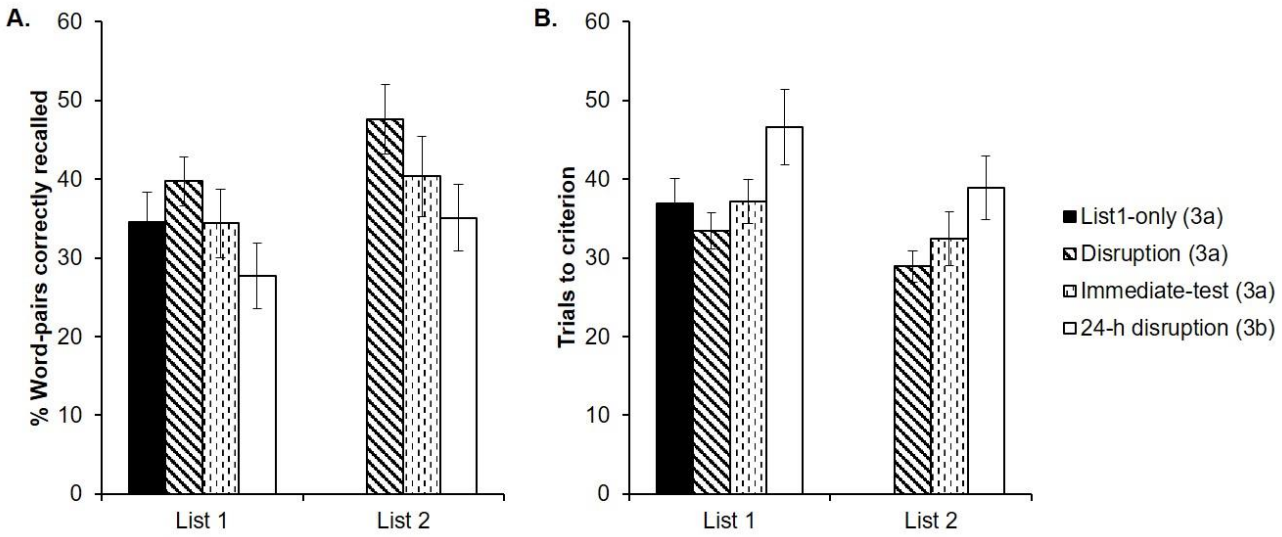

Figure 19. A. Mean number of trials to criterion for different groups in Experiment 3a and 3b (Session 1). B. Percentage of word-pairs correctly recalled after initial encoding. (Error bars: ± SEM).

Test 1, 2 and the amount forgotten

Participants in the 24-hour disruption group showed a considerable drop from Test 1 to Test 2. Furthermore, an ANOVA including all groups from Experiments 3a and 3b and the amount of forgetting as the dependent variable revealed a significant effect of group, $F(3, 75) = 3.899$, $p < 0.05$,

η_p^2 = 0.135 (Figure 20). Tukey's post-hoc tests showed that the 24-hour Disruption Group displayed significantly greater forgetting than the List-1-Only Group ($p < 0.01$, $d = 0.87$), but there was no statistically significant difference between the 24-hour Disruption Group and the Disruption Group ($p > 0.5$) from Experiment 3a or the Immediate-Test Group ($p > 0.1$), despite the 24-hour Disruption Group showing a greater mean amount of forgetting.

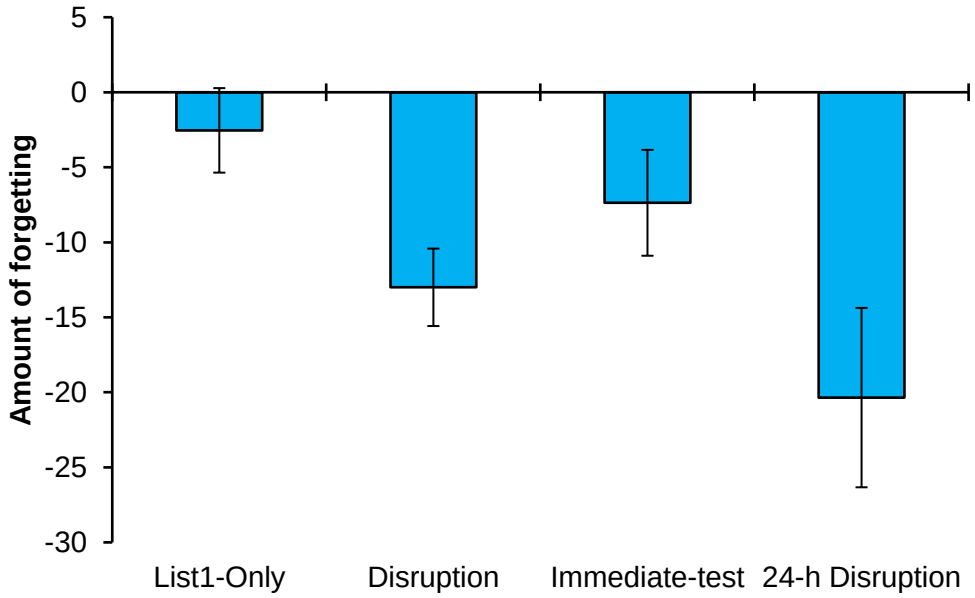

Figure 20. Mean amount of forgetting for the List1-Only (3a), Disruption (3a), Immediate-Test (3a) and 24-h Disruption groups (3b) (Forgetting = Session 2 score – Session 1 score). (Error bars: ± SEM).

Discussion

In Experiment 3b, I hypothesised that learning List 2 before retrieving List 1 would have little or no effect on memory performance because List 1 would have sufficient time to consolidate. Instead, when participants learned a second list of Swahili-English paired associates before being tested on an earlier list this produced significantly worse memory recall. This finding suggests that the lower recall of the disruption group in Experiment 3a may have been due to interference during retrieval instead of being related to memory consolidation. The influence of this interference may only manifest weakly when memory is tested immediately because the List 1 memory is still relatively strong but, after 24

hours, the List 1 memory may be weaker due to forgetting and consequently the interference from List 2 may be stronger and produce the memory deficit I observed.

What could be the source of this interference? One possibility is that while the Swahili stimulus terms are all different, there is probably some overlap in the sounds used and this interferes with participants recall. This would explain why Experiment 3a found an effect of learning a second set of paired associations and Experiment 2 did not. Many of the studies that use the A-B A-C paradigm use English words and, due to greater familiarity, the components that make up these words might be more tightly integrated into a single unit and no interference is observed (Bower, Thompson-Schill, & Tulving, 1994; Tulving & Watkins, 1974).

It is also interesting to note that the retroactive interference observed in Experiments 3a and 3b has somewhat different characteristics to the retroactive interference observed in past experiments. Previous experiments using the A-B A-C paradigm on a test soon after learning have found an initially high degree of retroactive interference with the A-C association being particularly strong, however, over time the A-C memory is partially forgotten and the A-B association partially recovers. In the experiment, it was observed that over time the retroactive deficit created by List 2 grew *larger*. This could simply reflect the influence of different kinds of retroactive interference. In the A-B A-C paradigm the interference might be produced from strongly interfering response terms, whereas in the current experiment the interference might be caused by weakly interfering stimulus terms.

There is, however, an alternative interpretation to the results of Experiment 3a. It is possible that the memory disruption observed in Experiment 3a was due to the memory for the initial list being labile and more susceptible to interference, while the memory disruption observed in Experiment 3b occurred because List 2 was being consolidated and was therefore still being actively processed by the brain. Because List 2 was being actively processed it may have disturbed retrieval of List 1 whereas if it was less active there would be no interference. One way of distinguishing between the different

interpretations of our results is by conducting an experiment where participants learn List 1 on day 1, List 2 on day 2 and are tested on List 1 on day 3. If the memory deficit is caused primarily by interference, performance should be worse for List 1 on day 3 for a group that learned List 2 compared to a group that did not learn List 2.

In experiment 3a, I found that learning second list of Swahili-English paired associates impairs the memory of a first list a small amount when tested immediately but substantially more when tested after a 24 hour delay. Initially, I believed this disruption of memory was most likely due to a disturbance of memory consolidation or the memory being in a labile state that made it more susceptible to interference. However, in Experiment 3b I found that when a second list is learned 24-hours later just before retrieving List 1 it creates an even greater disruption of memory recall. This means it is ambiguous whether learning a second list disturbs consolidation or memory retrieval and suggests that this methodology may not be ideal for exploring reconsolidation and the spacing effect.

3.5 Experiment 4a

As Experiments 1 to 3 denote, I planned to explore reconsolidation and the spacing effect with a word-based declarative memory task. Obviously, finding a methodology that unambiguously allowed us to explore the influence of variables on memory reconsolidation proved more difficult than I expected. One alternative to word-based tasks is the sequence tapping task. The task has been used by multiple research groups for investigating consolidation and reconsolidation (Censor, Dimyan, & Cohen, 2010; De Beukelaar et al., 2014; Karni et al., 1998; Walker et al., 2003). In this task, a sequence of numbers appears onscreen and participants have to type it as quickly and accurately as possible with their left hand. Consolidation and reconsolidation can be disrupted when participants learn a second sequence soon after learning or reactivating the first sequence and there is research to suggest it meets the criteria for an appropriate paradigm with which to test the reconsolidation account of the spacing effect. In particular, it has been demonstrated the memory disruption is time-dependent, such that if learning the second sequence is delayed it no longer disrupts the motor memory for the first sequence when tested 24 hours later (Walker et al., 2003). Additionally, after learning the second sequence, if an immediate test is conducted of the first sequence no memory deficit is observed (Walker et al., 2003).

In Experiment 4, I aimed to use the sequence tapping task to observe the reconsolidation process and the spacing effect under a similar set of circumstances. I used four groups of participants. The Spaced Group learned a sequence, completed a short repetition 24 hours later, and completed a retention test 1 week after the repetition session. The Massed Group learned a sequence, completed a short repetition immediately, and then completed a retention test 1 week later. Together, the Spaced and Massed groups allowed us to test for a spacing effect. To observe the memory reconsolidation process I compared the performance of a Disruption Group and a Control Group. The Disruption Group

learned a sequence, completed a repetition 24 hours later, and immediately after the repetition learned a second competing sequence. Finally, a memory test was conducted 24 hours after learning the new sequence. The Control Group learned a sequence, completed a short repetition 24 hours later, and then completed a memory test 24 hours after the repetition. I hypothesised that I would observe both a spacing effect and disruption of memory due to the reconsolidation process. That is, the Spaced Group would show better performance than the Massed Group and the Control Group would show better performance than the Disruption Group.

Method

Participants

Sixty-eight participants completed the experiment (18 male, 50 female; Age range: 18 – 32 years old). Participants were recruited via advertisements placed around the University of Otago and via the Department of Psychology student experiment participation system. Participants either received $20 compensation at the end of their third session or course credits.

Task

For the sequence-tapping task, participants sat at computer and were asked to prepare to type a sequence of numbers. They were asked to position their hand on the keyboard such that the little finger of their left hand was on the number 1 key, their ring finger was on the number 2 key, their middle finger on the number 3 key, and their index finger on the number 4 key. Participants were also informed that they needed to repeatedly type the sequence on the screen as quickly and accurately as possible. Each trial lasted 30 seconds and throughout the trial the sequence (either 4-1-3-2-4 or 2-3-1-4-2) was shown on the computer screen in large black letters on a white background. Participants were also provided with feedback throughout each trial: as each number of the sequence was typed a black dot appeared under the sequence. After all five numbers were typed the black dots disappeared, the sequence reset and participants began again at the start of the sequence. After each 30-second trial there

was a 30-second rest period. During the rest period on-screen instructions reminded participants to type as quickly and accurately as possible during the trial period. The task was presented using E-Prime 2 (Psychology Software Tools, 2013) on a standard desktop computer.

Procedure

As noted above, participants were randomly assigned to one of four groups. The training for all groups consisted of twelve 30 second trials with a 30 second rest period between trials. Participants in all groups were randomly assigned to use one of two sequences (4-1-3-2-4 and 2-3-1-4-2). These sequences have been used in a number of previous studies on consolidation and reconsolidation (e.g. De Beukelaar et al., 2014; Walker et al., 2003). All groups participated in three separate sessions. In Session 1, the Spaced Group completed 12 training trials. In Session 2, approximately 24-hours later, the Spaced Group completed a single repetition trial. In Session 3, approximately 1 week after Session 2, the Spaced Group completed a retention test consisting of three trials of the sequence they had learned. For the Massed Group, Session 1 consisted of a series of simple addition and subtraction equations. In Session 2, 24-hours later, the Massed Group completed both the training trials and the single repetition trial. Identical to the Spaced Group, Session 3 was approximately 1 week later and consisted of a retention test. For the Control Group, Session 1 consisted of 12 training trials, Session 2 (24 hours later) consisted of a single repetition trial, and Session 3 (24 hours after Session 2) consisted of a retention test. The Disruption Group's procedure was identical to the Control Group with the exception that, after the repetition trial in Session 2, they were trained on a second competing sequence for 12 trials.

Data Analysis

In tasks where speed and accuracy are components of participants performance a change in participants skill level is best represented by a shift in the speed-accuracy trade-off function (De Beukelaar et al., 2014). A speed-accuracy function (SAF) score was calculated for each trial. The

speed/inter-tap interval was calculated by averaging the time (in msec.) between each number typed and then dividing this number by 1000. Accuracy was calculated by dividing the number of correct numbers typed by the total numbers typed in a trial and then multiplying it by 100. The SAF score was then calculated by dividing the accuracy score by the speed score. To evaluate how well participants performed in Session 3, the SAF score on the first trial of Session 3 was divided by the SAF score of their repetition trial in Session 2. Additionally, I used planned contrasts and one-tailed tests to examine the spacing effect and disruption effect as I had strong expectations regarding the direction of the effects based on past research (e.g., Cepeda et al., 2006; De Beukelaar et al., 2014; Walker et al., 2003).

Results

Initial learning

Participant's performance learning the sequence tapping task is shown in Figure 21. There were no significant differences in performance between the groups on the average of the last three training trials (Trials 10, 11 and 12), $F(3, 76) = 1.281$, $p > 0.25$, $\eta_p^2 = 0.048$.

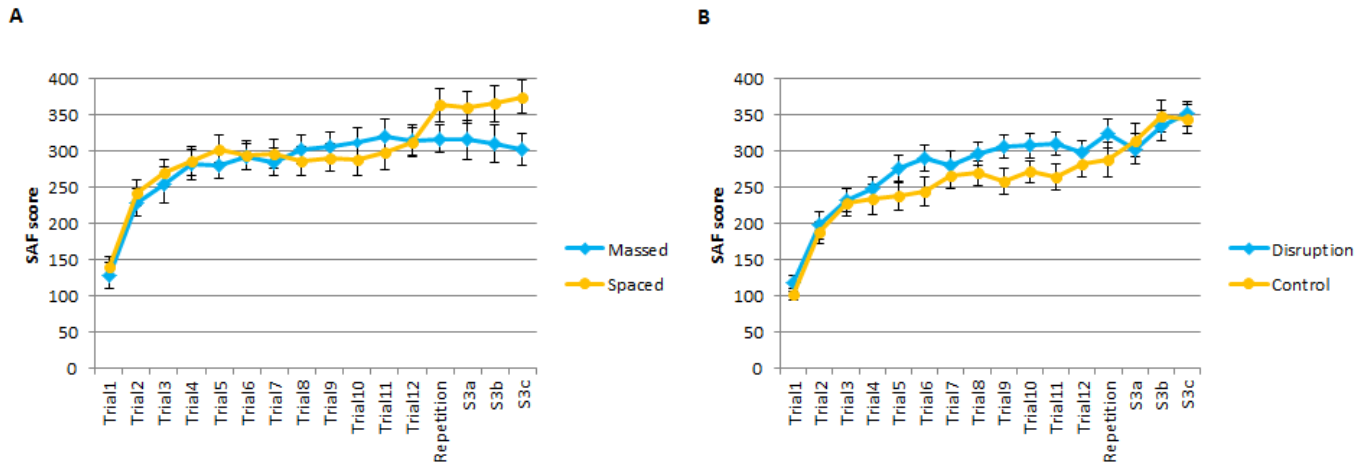

Figure 21. Experiment 4a sequence tapping experiment: **A**. SAF scores across trials for participants in the massed and spaced groups. **B.** SAF scores across trials for participants in the disruption and control groups. (Error bars: ± SEM).

Repetition trial

Past research has found that performance improves over a night of sleep on this task (De Beukelaar et al., 2014; Walker et al., 2003). I examined if an overnight improvement was present in the current study by collapsing across the Spaced, Control and Disruption groups and then conducting a paired t-test on the SAF score average of participant's last three training trials and the repetition trial. There was a significant improvement on participants repetition trial relative to the average on their last three training trials, $t(79) = 3.579$, $p < 0.001$, $d = 0.237$. I also examined whether the improvement seen in the groups whose repetition trial had a 24-hour delay was greater than the improvement in the Massed Group. A Mann-Whitney-U test was used because of significant violations of normality and large differences in sample size. The Overnight Delay Group showed a significantly greater improvement than the Massed Group, $U = 350.0$, $p < 0.01$, $g = 0.43$.

Final test

On the first trial of Session 3, although the Spaced Group was performing slightly better than the Massed Group (Figure 21), there was no significant effect of group, $F(3, 76) = 0.916$, $p > 0.4$, $\eta_p^2 = 0.035$. With respect to participant's performance on Session 3 relative to their performance on the Repetition Trial (S3a/rep), there was a significant effect of group, $F(3, 76) = 3.025$, $p = 0.035$, $\eta_p^2 = 0.107$. Planned contrasts revealed that the Control Group performed significantly better than the Disruption Group. There was, however, no significant difference between the Spaced and Massed groups.

Discussion

I hypothesised that the Spaced Group would perform better than the Massed Group and that the Control Group would perform better than the Disruption Group. These hypotheses were partially supported, the Control Group performed significantly better than the Disruption Group. Although the mean SAF score on Session 3 was slightly higher for the Spaced Group than the Massed Group, this

difference did not reach statistical significance. Part of the reason I wanted to observe the spacing effect and the reconsolidation process under the same set of circumstances was to assess whether the strength of training influenced the spacing effect in a similar manner to which it influences memory reconsolidation. My prediction is that decreasing the amount of training will lead to a larger effect size for the spacing effect. To answer this question, in Experiment 4b I set about manipulating the number of training trials participants received on the sequence tapping task.

3.6 Experiment 4b

The aim of Experiment 4b was to test one of the predictions of the reconsolidation account of the spacing effect. Specifically, if the reconsolidation account is correct, factors which influence memory reconsolidation should also influence the spacing effect. In the memory reconsolidation literature, one observation is that increasing the strength of a memory makes it *more* difficult for reconsolidation to occur (Suzuki et al., 2004; Wang et al., 2009; Winters, Tucci, & DaCosta-Furtado, 2009). Thus, if the reconsolidation account of the spacing effect holds, increasing the strength of a memory should therefore reduce the size of the spacing effect. To test this prediction, I manipulated the amount of training participants received on the sequence tapping task. Participants received either weak (6 trials) or strong (12 trials) training and then completed an additional trial after either 30 seconds (Massed Group) or 24 hours (Spaced Group). Both groups then completed a retention test 1 week later. Control and Disruption groups were also used to establish that the reconsolidation process was influenced by the strength of training in this task. Based on the reconsolidation account, I hypothesized that the participants who received weak training would show a larger effect size for the disruption and spacing effects than participants who received strong training.

Method

Participants

One hundred and eighty-one right handed participants took part in the current study (Mean Age = 22, range = 18 – 59). An additional 45 participants were excluded due to the fact they did not complete all three experimental sessions. One further participant that completed all three sessions was excluded because they failed to reliably improve on their trial 1 score. Participants were recruited via advertisements placed around the University of Otago and via the Department of Psychology student experiment participation system. Participants received $20 compensation at the end of their third

session. Participants were also screened so that they did not have extensive gaming or music experience and were right handed.

Task

The task was identical to that employed in Experiment 4a.

Procedure

Participant's sessions were generally scheduled for the same time each day, for the few for whom this was not possible, Session 2 and Session 3 were approximately 1 hour earlier or later than Session 1. There were eight groups in total, with half the participant in the Spaced, Massed, Control and Disruption groups receiving strong training and the other half receiving weak training. Participants that received strong training completed 12 training trials, 1 reactivation trial and 3 test trials. Participants receiving weak training completed 6 training trials, 1 reactivation trial and 3 test trials. Session 1 was identical for the Spaced, Control, and Disruption groups, with approximately half the participants in each group receiving either 12 (i.e., strong training) or 6 (i.e., weak training) trials. For participants in the Massed Group Session 1 consisted of a series of addition and subtraction equations.

For all groups, Session 2 was approximately 24-hours after the Session 1. In Session 2, the Spaced and Control groups completed one 30-second repetition trial and were then dismissed. The Disruption Group completed one 30-second repetition trial and then learned a second sequence for 12 trials. Finally, participants in the Massed Group learned the sequence across 12 (i.e., strong) or 6 (i.e., weak) trials and then completed their repetition trial after 30 seconds.

For the Spaced and Massed groups, Session 3 was 1 week after Session 2, while for the Control and Disruption groups it was 24 hours after Session 2. In Session 3, the Spaced, Massed, and Control groups completed three trials with rest periods in between and were then debriefed. For the Disruption Group, Session 3 consisted of them completing three trials with the sequence they originally learned and then three trials of the second sequence they learned.

Data analysis

I calculated a SAF score and participant's performance in their second session relative to the end of training (S2/S1) and their third session relative to their repetition (S3/rep) in the same manner as Experiment 4a. To test my main hypothesis that the strength of training would influence the size of the spacing and disruption effects I conducted two separate one-way ANOVA's, one for the strong training sub-group and one for the weak training sub-group. The ANOVA used the S3/rep score as the dependent variable and group (Spaced, Massed, Control, and Disruption) as the independent variable. Finally, I conducted planned contrasts that compared the Spaced group to the Massed group and the Control group to the Disruption group to identify precise group differences.

To examine whether improvement occurred as a function of sleep, I combined the Spaced, Disruption, and Control groups (i.e., the Overnight-Delay groups) and conducted separate paired sample t-tests for the strong training and weak training sub-groups. The Massed groups were not included in this analysis because their repetition trial occurred the same day as the original training. Further, explorations lead us to examine whether there was a difference in the improvement observed in the Overnight-Delay groups that completed strong or weak training. To make this comparison, I calculated an improvement score by dividing the SAF score on the reactivation trial by the average of the SAF score of the last three training trials. I then conducted a one-way ANOVA with training strength as the independent variable and the improvement score as the dependent variable. Finally, to examine whether the overnight-delay groups improved more than the Massed Group, I collapsed across the training-strength sub-groups and used a Mann-Whitney U test to examine differences in participant's improvement scores. The Mann-Whitney-U test was used because by combining the different groups there were significant violations of normality and large differences in sample sizes.

Results

Initial learning

I first examined whether the strength of training influenced the amount of learning achieved by comparing the SAF score on the repetition trial of participants who had received strong and weak training. Strong training (M = 314.71, SD = 97.83) lead to a significantly better SAF score than weak training (M = 265.19, SD = 84.44), $t(179) = 3.654$, $p < 0.001$, $d = 0.54$.

Session 3 performance

The data were approximately normally distributed, therefore, standard ANOVAs were employed for the initial analyses. I first examined whether there was an effect of group on participants SAF score on Session 3 relative to their reactivation trial in Session 2. For participants who received strong training there was no significant group effect, $F(3, 82) = 0.497$, $p > 0.65$, $\eta_p^2 = 0.018$. For participants who received weak training there was a significant effect of group and a larger effect size than for strong training, $F(3, 91) = 4.317$, $p < 0.01$, $\eta_p^2 = 0.125$. Planned contrasts for the weak training groups revealed that the Spaced Group (M = 1.09, SD = 0.15) retained significantly more skill than the Massed Group (M = 0.95, SD = 0.22), $t(91) = 2.690$, $p < 0.01$ (1-tailed) and the Control Group (M = 1.13, SD = 0.l6) performed significantly better than the Disruption Group (M = 1.04, SD = 0.18), $t(91) = -1.805$, $p < 0.05$ (1-tailed). To compare the individual effect sizes for the strong and weak training groups I also calculated Cohen's d. The effect size for spacing under strong training was $d = 0.15$ and the effect size of spacing under weak training was $d = 0.74$ (Figure 22). The effect size of the disruption effect under strong training was $d = 0.17$ and the effect size of disruption under weak training was $d = 0.52$ (Figure 22). These results suggest that weaker initial training produces larger spacing and disruption effects.

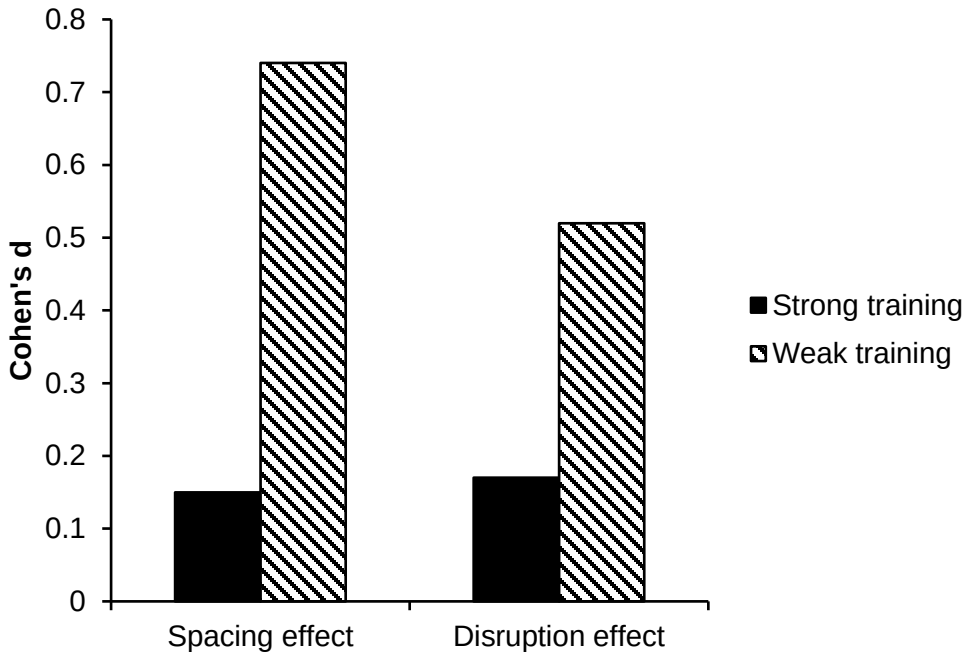

Figure 22. Effect sizes for the spacing and disruption effects under strong and weak training for Experiment 5b.

Overnight improvement

One alternative explanation for our results is a ceiling effect. That is, because strong training results in performance closer to participants plateau, there is less room for improvement and additional trials, whether spaced or massed, have little effect on the memory. Previous studies have reported an overnight improvement in scores on this task related to sleep (De Beukelaar et al., 2014; Walker et al., 2003) and this improvement allows us to test whether the strong training group had less opportunity to improve than the weak training group. I examined whether overnight improvement occurred in our data by combining the Spaced, Disruption and Control groups (i.e., the overnight delay groups). Participants who received strong and weak training showed significant overnight improvement (Strong: $t(65) = 3.531$, $p < 0.01$, $d = 0.249$ (1-tailed), Weak: $t(69) = 2.555$, $p < 0.01$, $d = 0.168$ (1-tailed)). I also tested whether there was a difference between the strong and weak groups in the size of

the overnight improvement and found no significant differences ($F(1, 179) = 0.906$, $p > 0.7$). To confirm that this improvement was due to the overnight delay rather than a general improvement, I compared the improvement of the Overnight-Delay groups to the improvement score of the Massed groups who had only a 30 second delay before their reactivation trial. The improvement of the overnight delay group was significantly greater than the massed group, U = 2555, $p < 0.05$ (1-tailed), $g = 0.2$.

Discussion

In the current study I tested predictions derived from a reconsolidation account of the spacing effect. I predicted that weaker training would produce a larger spacing effect and a larger disruption of the reconsolidation process. Consistent with these predictions, there was no statistically significant spacing effect or disruption effect when training was strong, but significant spacing and disruption effects occurred when training was weak. Furthermore, the Cohen's *d* and partial-eta square values confirmed that the spacing and disruption effects were larger when weak training was employed.

It is important to note there are a number of competing theories that aim to explain the spacing effect. One prominent group of theories are contextual variability theories (Estes, 1955; Glenberg, 1979; Pashler et al., 2009; Raaijmakers, 2003). These theories posit that a) spacing increases the contextual variability integrated into a memory during a repetition and b) additional variability makes retrieval more likely at a delayed retention test. If contextual variability theories are correct, in the current study I should have observed that the effect size of the spacing effect remains the same for different levels of training strength. Contextual variability theories make this prediction because the relative contextual variability of the spaced and massed groups should be approximately the same for strong and weak training. Given our findings, contextual variability theories are unlikely to explain the spacing effect observed.

Other accounts of the spacing effect suggests that it occurs because spacing makes memory retrieval more difficult and when a successful retrieval is more difficult it leads to a stronger memory

(Benjamin & Tullis, 2010; Bjork & Bjork, 1992; Delaney et al., 2010). When extra training is given retrieval difficulty theories predict a smaller effect size for the spacing effect because extra trials will make the memory easier to retrieve at the repetition. The main results of the current experiment are therefore consistent with retrieval difficulty theories. However, a close look at the data collected during the repetition trial suggests that the results of the current experiment are inconsistent with this theory. Previous studies have used speed and/or accuracy of responding as measures of retrieval difficulty (Kılıç, Hoyer, & Howard, 2013; Pyc & Rawson, 2009; Xue et al., 2011). In the current experiment I observed an improvement in participant's SAF scores after the overnight delay but no similar improvement in the Massed Group with a 30-second delay. This improvement means that spaced participants are able to retrieve the pattern of muscle movements faster and more accurately and, therefore, suggests that retrieval during the repetition trial was *easier* for spaced participants than massed participants. The current experiment along with other motor skill spacing studies (Dail & Christina, 2004; Shea et al., 2000) suggests that retrieval difficulty is unlikely to be the primary mechanism of the spacing effect over delays of 24 hours or more in motor skill studies.

Additionally, there is a curious difference between Experiment 4a and Experiment 4b. In Experiment 4a the partial-eta squared effect size of group was 0.104, whereas in Experiment 4b the strong training sub-group had a partial eta squared value of 0.018. The only procedural difference between these two groups was Experiment 4b used a laptop and desktop computer to run participants whereas Experiment 4a used only a desktop computer. Perhaps the laptop computer affected how participants typed or perceived the task. Alternatively, since I conducted this experiment, Hardwicke, Taqi, and Shanks (2016) published a failed replication of Walker et al.'s (2003) finding of a memory disruption in the sequence tapping task. Walker and Stickgold (2016) in their reply, noted that Hardwicke et al. (2016) used a variety of times to run participants whereas in their original study they had consistently used 1 pm to run the experiment and the time of day was known to influence

reconsolidation effects. While there were no intentional differences between the time of day which participants were run in Experiment 4a and 4b the time of day was largely based on when participants were available and perhaps uncontrolled variations in this variable is responsible for this difference between the two experiments.

In summary, the reconsolidation account of the spacing effect holds that the delay employed in spacing studies allows time for the memory to consolidate, resulting in the subsequent reconsolidation processes being more effective. In the current study, I tested this account by manipulating the strength of training and examining whether this had the same effect on the spacing effect and the reconsolidation process. Consistent with the reconsolidation account, I found that weak training lead to a greater spacing effect and a more sensitive reconsolidation process.

4 Chapter 4: Are there different mechanisms for the spacing effect across different timescales?

4.1 Experiment 5a

Memory reconsolidation takes time to develop. When a protein synthesis inhibitor or a behavioural intervention is provided during the memory reconsolidation process its effects are only detected many hours later. This is probably because there is a stable short-term memory but long-term memory depends on the reconsolidation process. As noted in the Introduction, the spacing effect is observed when short spacing intervals and retention delays of seconds and minutes are used as well as longer intervals of hours and days. Reconsolidation, therefore, is unlikely to explain the spacing effect over short timescales. This means if the reconsolidation account of the spacing effect is correct then different mechanisms must explain the spacing effect over different timescales.

One prominent explanation of the spacing effect is based on retrieval difficulty (Benjamin & Tullis, 2010; Bjork & Bjork, 1992). The authors of these theories suggest that due to the forgetting that occurs during a spacing interval, retrieval is more difficult and the repetition is processed more effectively. There is some evidence favourable to the retrieval difficulty theories. Bjork and Allen (1970) compared groups of participants who learned repeated trigrams with massed and spaced practice as well as manipulating the activity that filled the spacing interval. Some participants completed a difficult shadowing task whereas other participants completed an easy shadowing task. Bjork and Allen (1970) found that the more difficult activity (which presumably makes retrieval more difficult) enhanced recall. Pyc and Rawson (2009) also found evidence for the retrieval difficulty hypothesis by manipulating the spacing of items and the number of correct retrieval trials. Consistent with more difficult retrieval, spaced repetitions had longer response latencies than massed repetitions and a

decrease in response latency reflected the diminishing returns of additional repetitions for long-term retention.

Another piece of evidence relevant to retrieval difficulty theories comes from research examining the effects of testing memory. In general, a memory test benefits retention to a greater degree than restudying the same material (Cull, 2000; Hogan & Kintsch, 1971; Roediger III & Karpicke, 2006). Furthermore, free-recall tests result in better retention performance than cued-recall tests and cued-recall tests result in better retention performance than recognition tests (Bjork & Whitten, 1974; Carpenter & DeLosh, 2006; Glover, 1989). The relative benefit of each type of test is most likely due to differences in the difficulty of retrieval. Free-recall tests are most difficult because participants are given minimal cues to aid recall; they are simply told to recall all of the items learned in the experiment. Cued-recall tests are of intermediate difficulty because participants are given a cue to aid recall but they still need to generate the response based on the cue. Recognition tests are the easiest type of test because participants are provided with the item and merely need to remember whether they learned the item in the experiment or decide that it is a new item.

Based on the available evidence I believe that retrieval difficulty is the primary mechanism for the spacing effect over short timescales and memory reconsolidation is the primary mechanism for the spacing effect over long timescales. In Experiment 5a, I sought to test whether retrieval difficulty might be the mechanism that produces the spacing effect over short timescales. In the experiment I compared participant's performance in three conditions. In the Massed condition participants learned a paired-associate, completed an immediate cued-recall test, then studied the paired associate again and after a delay period completed a retention test. In the Spaced-Easy and Spaced-Hard condition, participants completed a study trial on a paired associate, studied other paired associates to form a spacing interval, completed a re-study trial and a test trial and then after a delay period completed a memory test. The only difference between the Spaced-Easy and Spaced-Hard condition was that the order of the trials after the spacing interval was manipulated to influence retrieval difficulty. In the Spaced-Easy

Condition participants completed a re-study trial and then a test trial, whereas in the Spaced-Hard Condition participants completed a test trial and then a re-study trial. I hypothesised that the Spaced-Hard group would show better retention performance than the Spaced-Easy Condition and that the Spaced-Easy Condition would show better retention performance than the Massed Condition.

Method

Participants

Eighty participants were recruited through Mechanical Turk (https://www.mturk.com/). All participants lived in the United States (Age range: 22 – 57, Mean Age = 35). There were 21 females and 57 males, with 2 participants not filling in demographic information. Participants were paid $8 US for completing the experiment. Participants were told the experiment was aimed at understanding the relationships between repetition and memory.

Materials

All participants completed the experiment on their personal computers. The experiment was created with Inquisit Lab and presented using Inquisit Web software (Millisecond Software: https://www.millisecond.com/). Three lists were created by randomly selecting Indonesian-English word pairs from a list downloaded from https://sites.williams.edu/nk2/stimuli/. Each list consisted of 15 Indonesian-English word-pairs.

Procedure

The experimental design was a within-subjects design such that each participant completed all three conditions. Participants began the experiment by following a link which enabled them to install the software necessary for Inquisit web and click an additional link to begin the experiment. Participants were randomly assigned to one of the eighteen experimental orders. Participants filled in their demographic details and were given instructions for the experimental session. Participants were told that they would be studying and recalling a series of Indonesian-English word-pairs and to study

the pairs so that when given the Indonesian word they can recall the English equivalent. Additionally, they were told that when a word-pair is repeated to do their best to strengthen the initial connection they have made.

For each condition participants there was a learning phase, 30 seconds of simple math equations, and a test phase. During the learning phase each of the 15 word-pairs were presented for three trials, two study trials and one test trial. In a study trial the word-pair was presented for 5 seconds. In a test trial, the Indonesian word was presented as a cue and participants were asked to type the English equivalent. If participants did not complete a response within 10 seconds the trial automatically ended. The test phase simply consisted of one test trial for each word-pair and was identical across the different conditions. The learning phase, however, differed for each condition.

In the Massed Condition, the learning phase consisted of presenting participants with a study trial for a word-pair followed immediately by a test trial for that word-pair and then a second study trial. The lag in this condition was 0 trials. Following this, participants completed simple math equations and then the test phase.

In the Spaced-Easy Condition participants studied the first five word-pairs of the list. They then completed a second study trial on the first word-pair presented. Thus, the first word-pair was presented with a spacing interval of approximately 20 seconds or a lag of 4 trials. Next, participants completed a test trial on the first word-pair. After the test trial the participants completed a second study trial on the second word-pair presented and then a test trial on the second word-pair. This procedure continued until all five word-pairs and had been repeated. Next, participants studied the second five word-pairs of the list and then repeated the word-pairs with a study and test trial in the same manner as the first five word-pairs. Participants then studied the final five word-pairs of the list and repeated each word-pair as before with one study trial and then one test trial. The average lag in this condition was 6 trials. Participants then completed 30 seconds of math equations and the test phase.

In the Spaced-Hard condition participants studied the first five word-pairs of the list. They then completed a test trial of the first word-pair presented. After the test trial participants completed another study trial of the first word-pair. Next participants completed a test trial and then a study trial of the second word-pair and so on until the first five word pairs of the list were all repeated with a test trial and a study trial. Participants then studied the second five word-pairs of the list and repeated them with a test trial and study trial. Next participants completed the learning phase by studying the final five word-pairs and then repeating them with a test trial and a study trial. The average lag in this condition was 6 trials. After the learning phase participants completed 30 seconds of math equations and the test phase.

In the test phase participants completed a test trial on all 15 word-pairs. The test trial presented participants with the Indonesian word as a cue and they had to type its English equivalent. No feedback was given during the test. After participants had completed all three conditions they were presented with a completion code. They then sent the completion code to the experimenter and were given compensation.

Results

Learning phase performance

I first analysed participant's accuracy on the test trial during the learning phase using a repeated measures ANOVA. There was a significant effect of condition on the percentage correct on test trials during the learning phase $F(2, 158) = 115.90$, $p < 0.001$, $\eta_p^2 = 0.663$. Consistent with differences in retrieval difficulty, within the Spaced-Easy Condition participants showed a higher percentage of words recalled correctly on test trials than within the Spaced-Hard Condition ($p < 0.0001$, $d = 1.53$) (Figure 23A). Additionally, the Spaced-Easy condition was not significantly different from the Massed Condtion, ($p > 0.3$).

I also analysed participant's latency to type a response during test trials. There was a significant effect of group, $F(2, 158) = 64.06$, $p < 0.001$, $\eta_p^2 = 0.448$. Again consistent with differences in retrieval difficulty, within the Spaced-Easy condition participants showed significantly lower response latencies than within the Spaced-Hard Condition, $p < 0.0001$, $d = 1.11$ (Figure 23B). Additionally, in the Spaced-Easy Condition participants showed a significantly lower response latencies than in the Massed Condition, $p < 0.05$, $d = 0.2$.

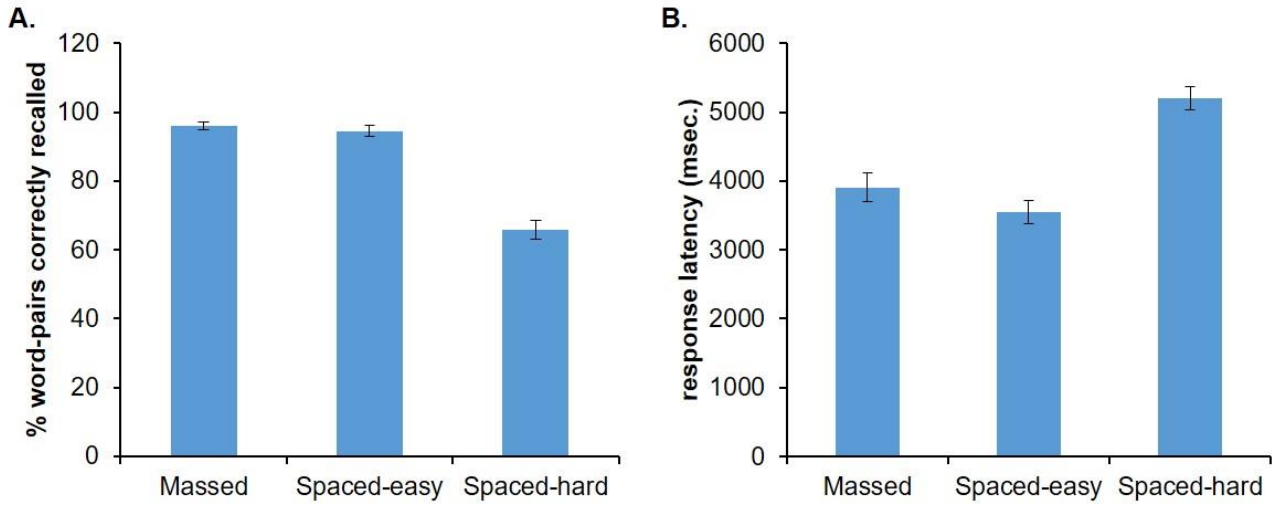

Figure 23. **A.** Mean percentage of word-pairs correctly recalled on the test trials during the learning phase. **B.** Mean response latencies on test trials during the learning phase (Experiment 5a). (Error bars: ± SEM).

Recall performance

I analysed the percentage of correctly recalled word-pairs on the final test using a repeated measures ANOVA. There was a significant effect of condition on percentage of correctly recalled word-pairs $F(2, 158) = 115.90$, $p < 0.001$, $\eta_p^2 = 0.342$ (Figure 24). Within the Spaced-Easy Condition participants recalled a significantly lower percentage of word-pairs correctly than within the Spaced-Hard, $p < 0.005$, $d = 0.28$. Additionally, within the Spaced-Easy Condition participants recalled a significantly greater percentage of word-pairs correctly than in the Massed Condition, $p < 0.001$, $d = 0.31$.

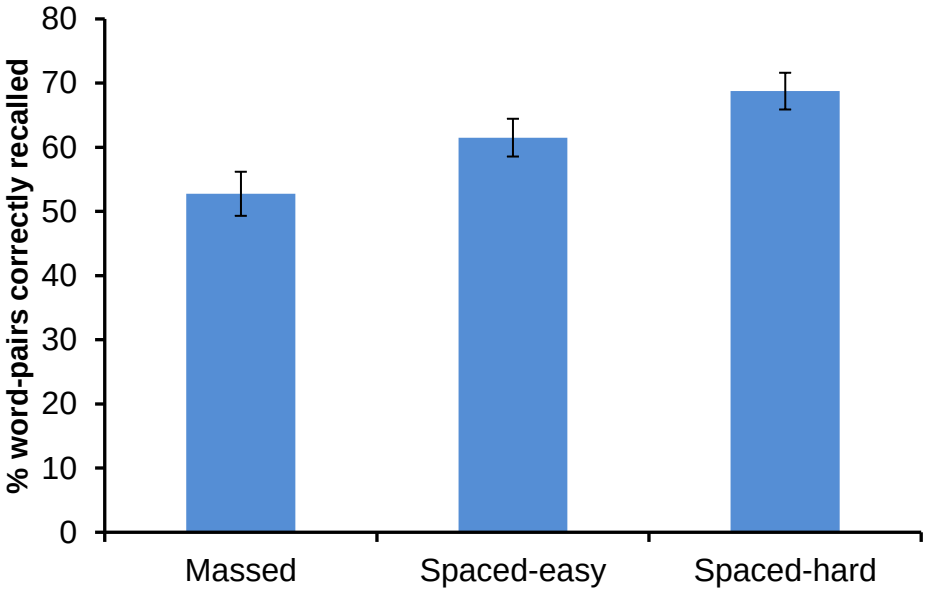

Figure 24. Mean percentage of word-pairs correctly recalled on the final recall test of Experiment 5a. (Error bars: ± SEM).

Discussion

I hypothesised that I would observe a spacing effect and that more difficult retrieval would produce better memory. The results of Experiment 5a were consistent with my hypotheses. Both the Spaced-Easy and the Spaced-Hard conditions showed better memory recall than the Massed Condition and the Spaced-Hard group showed better recall then the Spaced-Easy group. Additionally, consistent with the idea that the Spaced-Hard Condition had more difficult retrieval than the Spaced-Easy Condition, the latency and accuracy of the Spaced-Hard Condition was longer than the Spaced-Easy Condition and the accuracy for the test trials during the learning phase were less than the Spaced-Easy Condition.

4.2 Experiment 5b

In Experiment 5a, I found evidence consistent with the retrieval difficulty explanation of the spacing effect. In Experiment 5b I aimed to determine if manipulating retrieval difficulty has the same effect over long timescales as it does over small timescales. To make this determination, in Experiment 5b I examined the retention of paired-associates after 1 week with a spacing interval of 24 hours. Retrieval difficulty theories assume that retrieval difficulty influences the spacing effect over short and long timescales in a similar manner, so if these theories are correct I should observe a result similar to Experiment 5a. However, a key prediction of the reconsolidation account is that reconsolidation produces the spacing effect over long timescales and another mechanism produces the spacing effect over short timescales. Therefore, if the reconsolidation account is correct, the Spaced-Easy Group and Spaced-Hard Group should show a similar level of retention. I hypothesised that the Spaced-Easy and Spaced-Hard groups would show less forgetting than the Massed Group and, further, that the Spaced-Easy Group would not show greater forgetting than the Spaced-Hard Group.

Method

Participants

One hundred and sixty-five participants from Mechanical Turk completed this experiment. Of the 165 participants there were 65 females and 98 males and two participants who did not report any demographic information. All of the participants who reported demographic information were from the United States except one who was from the United Kingdom. The age range of participants was 19 – 62 with a mean age of 35. Participants were paid $3 USD for completing the first session, $6 USD for completing the second session, and $12 USD for completing the third session.

Materials

The experiment was created using Inquisit lab and presented to participants using Inquisit web. Participants in the experiment learned and were tested on 20 Swahili-English word-pairs randomly

chosen from the word-pairs normed by Nelson and Dunlosky (1994). The order in which the word-pairs were presented and tested throughout the experiment was randomised by the Inquisit software. Participants were also given a recognition test where the twenty previously learned words were mixed with twenty new words randomly chosen from Nelson and Dunlosky (1994). In the recognition test participants were asked to press "Y" for yes if they had previous seen the word-pair in the experiment and "N" for no if they had not previously seen the word-pair.

Procedure

All participants included in the experiment completed three sessions. Session 2 was completed approximately 24 hours after the Session 1 and Session 3 was completed 1 week after the Session 2. At the beginning of each session participants entered their session number and condition number. The condition number was randomly assigned at the beginning of the first session. In the first session all participants were asked to provide demographic information (country, city, age and sex) and then started the experiment.

In Session 1, participants in the Spaced-Easy Group and the Spaced-Hard Group learned the twenty Swahili-English word-pairs through a series of study and test trials. Participants were initially instructed to study the word-pairs so they could be remembered for a future test. They then completed twenty study trials, where each word-pair was presented for 5 seconds. Next participants completed a cued-recall test on each word (with no feedback being provided), with the Swahili word being presented as cue and participants being asked to type the English equivalent. Participants were given 8 seconds to type their response, however, if participants typed the word before 8 seconds elapsed they could move to the next pair by pressing the enter key or clicking the next button that was presented onscreen. After the initial cued-recall test participants were presented with additional study trials for the word-pairs they had not been able to recall correctly. They were then presented with additional cued-recall test trials for those word-pairs that had not been correctly recalled. This cycle of study and

test trials continued until all of the word-pairs were correctly recalled. Next participants completed an immediate test where each word-pair was tested again through cued-recall without feedback, the only difference from the previous test trials being participants were given 15 seconds to type their response.

In Session 2 participants in the Spaced-Easy Group began their session with a recognition test on the twenty Swahili-English word-pairs they had studied on the previous session. They then completed two repetitions consisting of two cued-recall tests on each word-pair with feedback being provided in the form of the correct word-pair being displayed for 3 seconds for both correct and incorrect responses. All twenty word-pairs were presented in the cued-recall before any word-pairs were repeated. The procedure for the Spaced-Hard Group was the same as the Spaced-Easy Group except that they completed the two cued-recall tests and then the recognition test.

For participants in the Massed Group, Session 1 consisted of maths equations for 60 seconds. In Session 2, participants were asked to learn twenty Swahili-English word-pairs and completed the same study-test cycle as the Spaced-Easy and Spaced-Hard groups completed in Session 1. After learning the word-pairs the massed group completed two cued-recall tests and then the recognition test.

In Session 3, all three groups completed a retention test. The retention test was simply another cued-recall test on all of the Swahili-English word-pairs learned in the previous sessions. Participants were given 15 second to type their response before the software tested the next word-pair.

Results

Learning phase performance

I examined whether the groups differed in the number of trials required to reach the criterion of recalling each word-pair correctly. There was no significant effect of group, $F(2, 162) = 1.332$, $p > 0.2$, despite the mean number of trials the Spaced-Easy Group required ($M = 29.65$, $SD = 33.06$), being slightly higher than the Spaced-Hard Group ($M = 22.67$, $SD = 16.75$, $d = 0.27$), and the Massed Group ($M = 24.43$, $SD = 15.83$, $d = 0.21$).

Manipulation check: cued-recall test 1

I was initially interested in whether there were differences in the latency and accuracy of recall in cued-recall test 1 consistent with a successful manipulation of retrieval difficulty. There was a significant group effect on latency scores, $F(2, 162) = 4.603$, $p < 0.05$, $\eta_p^2 = 0.054$ (Figure 25A). Consistent with an successful manipulation of retrieval difficulty the latency of the Spaced-Easy Group was significantly lower than the Spaced-Hard Group, $t(162) = -2.312$, $p < 0.005$, $d = 0.54$. Additionally, there was no significant difference in latency between the Massed Group and the Spaced-Easy Group, $t(162) = -0.372$, $p > 0.7$, $d = 0.069$. There was also a significant effect of group on accuracy for cued-recall test 1, $F(2, 162) = 12.925$, $p < 0.001$, $\eta_p^2 = 0.138$ (Figure 25B). Again consistent with a successful manipulation of retrieval difficulty, the Spaced-Easy Group showed more accurate recall than the Spaced-Hard Group, $t(162) = 2.550$, $p < 0.01$, $d = 0.45$. Additionally, the accuracy of the Massed Group was significantly higher than the Spaced-Easy Group, $t(162) = 2.276$, $p < 0.05$, $d = 0.47$.

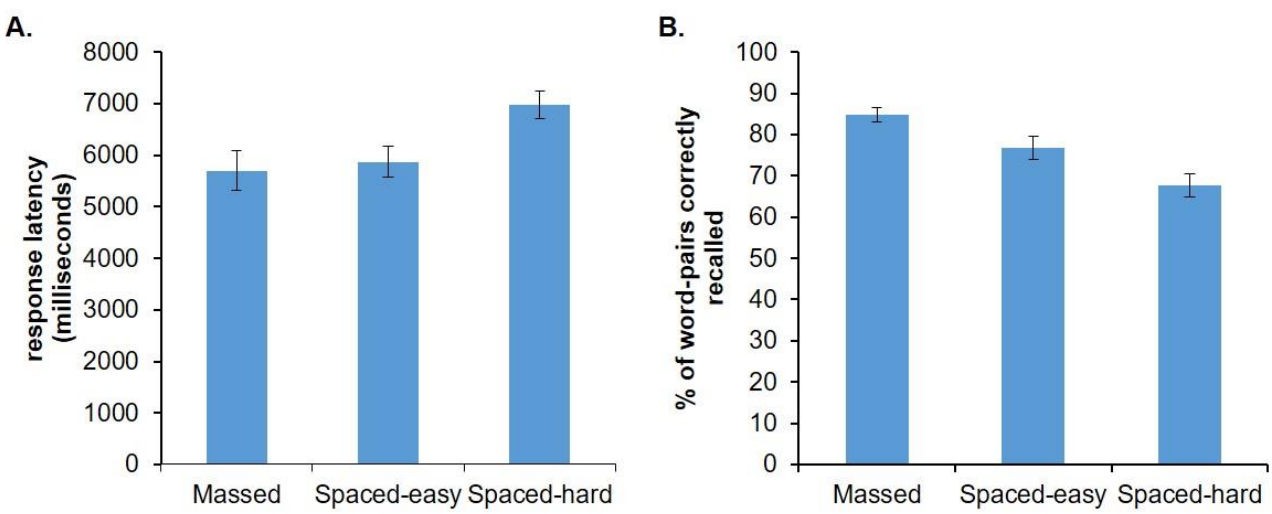

Figure 25. **A.** Mean response latency on the first cued-recall test. **B.** Mean percentage recall on the first cued-recall test (Experiment 5b, Session 2).

Cued-recall test 2

I also assessed whether there were group differences in cued-recall test 2. There was no significant group effect on latency, $F(2, 162) = 0.631$, $p = 0.533$, $\eta_p^2 = 0.08$. There was however a significant group effect for accuracy, $F(2, 162) = 5.33$, $p < 0.05$, $\eta_p^2 = 0.044$. The accuracy of the Massed Group ($M = 92.92$, $SD = 11.87$) was significantly higher than the Spaced-Easy Group ($M = 86.35$, $SD = 17.86$), $t(162) = 2.347$, $p < 0.05$, $d = 0.43$. There was no significant difference between the Spaced-Easy Group and the Spaced-Hard Group (M = 86.67, SD = 13.67) for accuracy, $t(162) = -0.110$, $p > 0.9$, $d = 0.02$.

Recognition test

We examined whether the different procedures of the groups influenced performance on the recognition test. There was a significant effect of group on the percentage of correct responses on the recognition test $F(2, 162) = 6.111$, $p < 0.01$, $\eta_p^2 = 0.071$. Performance on the recognition test reflects whether it comes before (Spaced-Easy) or after the cued-recall tests (Massed and Spaced-Hard). The Spaced-Easy Group ($M = 97.34$, $SD = 4.65$) showed a significantly lower mean percentage of correct responses than the Massed Group (M = 98.73, SD = 2.34), $p < 0.05$, $d = 0.38$ and the Spaced-Hard Group (M = 99.38, $SD = 1.37$), $p < 0.05$, $d = 0.60$.

Retention test performance and the amount of forgetting

We found a significant effect of group for the percentage of correctly recalled word-pairs on the retention test $F(2, 162) = 6.111$, $p < 0.05$, $\eta_p^2 = 0.037$. There was no significant difference between the Spaced-Easy Group ($M = 59.89$, $SD = 26.58$) and the Spaced-Hard Group (M = 61.75, $SD = 24.81$), $p > 0.7$, $d = 0.07$. When collapsed together ($M = 60.90$, $SD = 25.53$), however, the spaced groups showed significantly higher percentage recall than the Massed Group ($M = 50.58$, $SD = 26.63$), $p < 0.001$, $d = 0.4$.

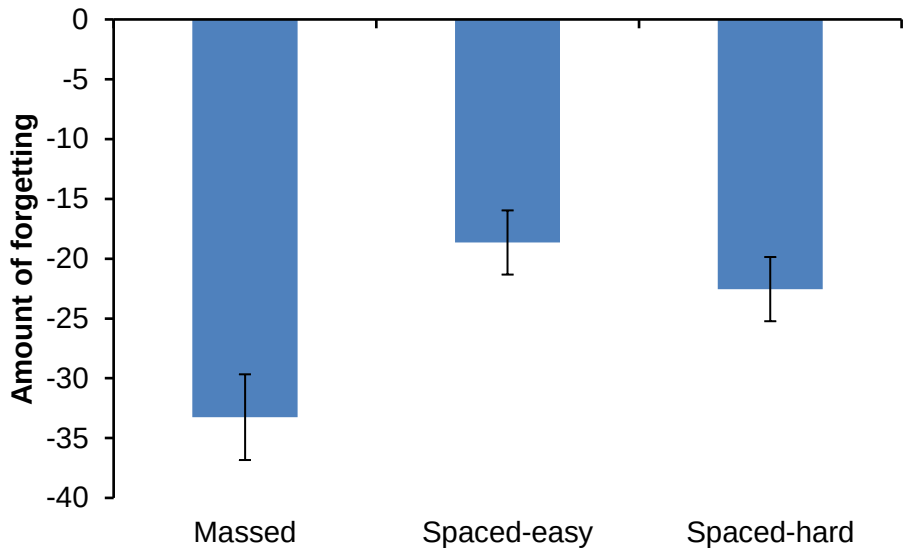

Figure 26. Mean amount of forgetting of the Massed, Spaced-Easy, and Spaced-Hard groups in Experiment 5b. (Error bars: ± SEM).

We also analysed the amount of forgetting of the different groups. There was a significant effect of group on the amount of forgetting, $F(2, 162) = 6.074$, $p < 0.005$, $\eta_p^2 = 0.07$ (Figure 26). I found that the combined spaced groups showed significantly more forgetting than the Massed Group, $t(162) = 3.416$, $p < 0.001$, $d = 0.52$. Additionally, although the Spaced-Easy Group forgot slightly less than the Spaced-Hard Group, there was no statistically significant difference in the forgetting of these two groups, $t(162) = 0.870$, $p > 0.3$, $d = 0.2$. Since our hypothesis was that the Spaced-Easy Group would not forget more than the Spaced-Hard Group, we also conducted a non-inferiority t-test to provide a more precise assessment of the data. In the non-inferiority t-test the null-hypothesis is that the Spaced-Easy Group shows greater forgetting than the Spaced-Hard Group (i.e., the Spaced-Easy Group is inferior to the Spaced-Hard Group) and the alternate hypothesis is that the Spaced-Easy Group does not show greater forgetting than the Spaced-Hard Group (i.e., the Spaced-Easy Group is not inferior to the Spaced-Hard Group). We found further evidence that the Spaced-Easy Group did not forget more

than the Spaced-Hard Group, with the non-inferiority test showing a statistically significant result $t(103) = 1.795$, $p < 0.05$.

Discussion

Based on the reconsolidation account of the spacing effect, I hypothesised that the Spaced-Easy Group would not perform worse than the Spaced-Hard Group. My results were consistent with this hypothesis. I observed that the Spaced-Easy Group did not forget more or have lower recall on the retention test than the Spaced-Hard Group. Additionally, I observed a spacing effect with the spaced groups performing better than the Massed Group. Combined with Experiment 5a my results support the prediction that different mechanisms produce the spacing effect over different timescales.

One aspect of retrieval difficulty theories is that while more difficult retrieval is beneficial, if retrieval is unsuccessful no benefit to memory will be obtained. At the time point of cued-recall test 1 the Spaced-Hard Group does perform significantly worse than the Spaced-Easy Group. Could my results therefore be due to greater retrieval failure diminishing the benefits of more difficult retrieval? I think this is a poor explanation of our results. Firstly, the performance of the Spaced-Hard Group on the second cued-recall test is almost identical to the Spaced-Easy Group (86.67% vs 86.35%). Secondly, other studies using short timescales have found greater retrieval difficulty beneficial despite differences in the number of correct responses. For example, Carpenter and DeLosh (2006) found that an intervening cued-recall test produced better final recall than a recognition test despite the mean proportion correct on the cued-recall test (M = 0.72) being significantly lower than the mean proportion correct on the recognition test ($M = 0.89$). Furthermore, in Experiment 5a the recall of the Spaced-Hard Group was significantly lower than the Spaced-Easy Group on the test trials during the learning phase, yet greater retrieval difficulty enhanced recall on the final test.

In our experiments we used response latency and accuracy scores for the cued-recall test as a measure of retrieval difficulty. One observation about Experiment 5b is that the effect size of the difference between the Spaced-Easy and Spaced-Hard groups is much smaller than in Experiment 5a

(response latency: d = 0.54 vs d = 1.11, accuracy: d = 0.45 vs d = 1.53). Could this difference in the strength of the retrieval difficulty manipulation explain the different results of the two experiments? If I had used only traditional statistical tests and a small sample size, then a finding of no statistically significant difference between the Spaced-Easy and Spaced-Hard groups could be due to differences in the influence of our manipulations. However, because I used a large sample size, a non-inferiority test, and our manipulation produced a statistically significant effect on retrieval difficulty, it is unlikely that the different results of the two experiments are due to differences in the strength of the retrieval difficulty manipulation.

The results of this experiment have implications for theories of the spacing effect. Theories of the spacing effect should be able to explain why retrieval difficulty seems to have a large influence over short time scales and little influence on the spacing effect when longer timescales are used. Retrieval difficulty theories and contextual variability theories as they are currently formulated assume the continuity of their mechanisms across short and long timescales, whereas this appears not to be the case. Part of the reason for the lack of continuity may be that there are differences between short and long-term memory. My results may reflect that for short-term memory differences in the quantity or quality of processing may produce a large difference in subsequent recall, whereas for long-term memory that has been modified by memory consolidation such differences in memory processing may have relatively little impact.

5 Chapter 5: General Discussion

The main purpose of this thesis was to develop a better understanding of the spacing effect. To this end, I studied the literature and developed a reconsolidation account of the spacing effect. I tested two different predictions based on the reconsolidation account. The first prediction was that as the strength of a memory increases it should influence reconsolidation and the spacing effect in a similar manner. After completing a number of experiments to obtain the right paradigm I tested and confirmed this prediction in Experiment 4b. A second prediction was that there should be evidence of different mechanisms across different timescales. I tested this prediction in Experiment 5a and 5b and found that retrieval difficulty appears to influence the spacing effect over short timescales but has little influence over long timescales.

One aspect of the current thesis worth reflecting on is why it took Experiments 1 – 4 to develop an appropriate methodology to test a reconsolidation account of the spacing effect. In Experiment 1, I failed to obtain a spacing effect and replicate the disruption of memory due to emotional faces obtained by Strange et al. (2010). The failure to obtain a spacing effect was most likely due to one of the boundary conditions of the spacing effect. If too much forgetting occurs during the spacing interval, a spacing effect will not be obtained. The failure to obtain the memory disruption could be due to differences in minor aspects of the experiment or possibly Strange et al.'s results are not reliable enough to be reproduced by independent researchers. Whatever reason produced our failure to replicate Strange et al. (2010), the solution is the same. Strange et al. (2010), and researchers in general, should make the materials of their experiments publicly available so other researchers can replicate their results precisely and with ease. Replication of past results is one area that psychological science needs to improve in general.

In Experiment 2, I discovered some researchers need to be more sceptical that memory deficits produced are due to a disturbance or consolidation or reconsolidation. Researchers had previously

found that a motor skill task following a free-recall task resulted in poorer memory performance 24 hours later and vice versa (Brown & Robertson, 2007). They attributed this drop in performance as being due to a disturbance of the consolidation of motor and declarative memories. However, our finding that two similar declarative tasks did not produce any disruption of reconsolidation suggests that consolidation and reconsolidation is highly specific. In some studies researchers have probably misinterpreted interference produced by overlapping contextual cues or strategy as a disruption of memory consolidation.

In Experiment 3a, I found that learning a second set of Swahili-English paired-associates produced a non-significant disruption of memory on an immediate test and produced a significant disruption of memory the following day. To confirm that this was a disruption of consolidation in Experiment 3b I tested the effect of learning the second set of paired-associates before retrieving the first set 24 hours after the initial learning session. Learning the second set seemed to disrupt retrieval of the first set and suggested that this was not a reliable method for determining the effect of variables on consolidation and reconsolidation. Future research on consolidation and reconsolidation should use this testing method to confirm that the memory deficit they observe is due to consolidation and not interference.

In the time I conducted these experiments there have been a number of attempts to replicate reconsolidation-related effects in humans. The basic findings for the sequence tapping task has been replicated in a number of studies (Censor et al., 2010; De Beukelaar, Woolley, Alaerts, Swinnen, & Wenderoth, 2016; De Beukelaar et al., 2014), however, some researchers have had difficulty replicating the disruption of fear memory during the reconsolidation process (Golkar, Bellander, Olsson, & Öhman, 2012; Kindt & Soeter, 2013) and one researcher did not replicate Walker et al.'s (2003) results in the sequence tapping task (Hardwicke et al., 2016). These results probably reflect a general principle of human memory research that all memory effects or phenomena depend on a

particular set of control variables (Roediger, 2008). For example, it was initially believed that the depth of processing effect where getting participants to encode stimuli in a meaningful way resulted in better memory than other "shallower" methods of processing was akin to a law of memory (Roediger, 2008). However, the depth of processing effect was later found to depend on several control variables such as the type of test and the type of materials used. In terms of memory reconsolidation, researcher's ability to observe the reconsolidation process in humans (with many uncontrolled variables) is probably influenced by a number of factors such as the type of materials, the instructions used, how easily participants forget the learning materials, and the degree of proactive and retroactive interference outside of the experiment.

In Experiment 4a, I found a disruption of reconsolidation using the sequence tapping task. Part of the reason for the success of this methodology was probably that it had been used effectively by multiple independent researchers for consolidation and reconsolidation research (Censor et al., 2010; De Beukelaar et al., 2014; Fischer & Born, 2009; Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Walker et al., 2003). Additionally, in past research this methodology had several checks to confirm that the additional sequence was disrupting consolidation and reconsolidation. Furthermore, using this task I found that the Spaced Group showed better performance than the Massed Group, although this difference was not statistically significant. These observations indicated that the sequence tapping task was a useful method for exploring whether variables that influence memory reconsolidation also influence the spacing effect in a similar manner.

In Experiment 4b, using the sequence tapping task, I manipulated the strength of training. Some participants received strong training (12 trials) and others received weak training (6 trials). Training strength influenced both reconsolidation and the spacing effect. Consistent with past research, the memory reconsolidation process was engaged less when strong initial training was used. Furthermore, with reconsolidation less engaged, strong training also led to a smaller effect size for the spacing effect.

This result is consistent with the spacing effect over long timescales being produced by memory reconsolidation. It is inconsistent with contextual variability theories of the spacing effect.

In Experiment 5a and 5b, I tested another prediction of the reconsolidation account of the spacing effect. The reconsolidation account predicted that I should observe discontinuity between the spacing effect over short and long timescales. In contrast, contextual variability and retrieval difficulty theories suggest that the same mechanism explains the spacing effect over both short and long timescales, therefore factors which are important over short timescales should also be important for the spacing effect over long timescales. In my experiment I observed that retrieval difficulty was important for short timescales, spaced repetitions where retrieval was difficult resulted in better retention than spaced repetitions where retrieval was easy. However, across long timescales spaced repetitions where retrieval was difficult did not produce better retention than spaced repetitions where retrieval was easy. There is therefore, a discontinuity in the influence of retrieval difficulty on the spacing effect over short and long timescales.

I reported in Chapter 1 that there is a similar inverted-U curve between spacing and retention for both short and long timescales. If different mechanisms explain the spacing effect over different timescales, why is there a similar inverted-U curve? The answer is that the inverted-U curve over both long and short timescales reflects a balance between the memorial benefits obtained from spacing and the costs of forgetting. Both reconsolidation and retrieval difficulty are automatic processes that provide a consistent benefit to memory and the amount of forgetting is fairly consistent across short and long timescales (Wixted & Ebbesen, 1991). The similarity which the two mechanisms function in this context results in similar inverted-U curves across both short and long timescales.

As discussed in Chapter 1 the reconsolidation account can explain some of the interesting patterns in the spacing literature using long timescales. Most of these patterns can be explained in terms of how memory reconsolidation interacts with prior participant's prior experience. The observation

that for skill-related tasks adults show enhanced learning and retention from spaced repetitions whereas for language-related tasks adults shown enhanced retention can potentially be explained by adults greater prior experience with language. Due to their experience adults can rapidly learn and recall words and therefore do not require consolidation and reconsolidation to reach a high level of performance. In contrast, for most skill-related tasks time between repetitions is necessary so consolidation and reconsolidation to make changes to acquire further skill. Similarly, the observation that adults generally do not show a spacing effect for learning in language-related tasks while children do, can be explained in-terms of children benefitting from consolidation and reconsolidation to enhance their acquisition of words, whereas for adults their greater prior experience means consolidation and reconsolidation do not enhance learning.

The balance between the benefits of reconsolidation, the influence of prior experience and forgetting can also provide a potential explanation for the effect of different spacing schedules on different tasks. We observed that Young (1954) found that a 2 days per week was better for practicing badminton, whereas 4 days per week was better for practicing archery. For archery it might be the case that more concentrated sessions are required to build up the basic skills with minimal forgetting and using the schedule of 2 days per week participants have significant forgetting between sessions and this undermines the benefits of reconsolidation from greater spacing. In contrast for badminton, participants probably have prior experience with a racket sport or hitting a ball in general (e.g., cricket, tennis, table tennis) and since the basic skills are established participants might show minimal forgetting between sessions and benefit from greater spacing to enhance reconsolidation. These same principles can also explain why Harmon and Miller (1950) found that an a schedule with gradually expanding spacing intervals was better than many other spacing schedules.

An additional question is: how does spacing and memory reconsolidation enhance the generalization of learning? The following account is tentative but is a worthwhile framework for future research. From what we know about consolidation it seems likely that consolidation can function to

help extract the general features of a task. Having time between repetitions allows the brain to progress more in forming a general framework for a task and having a general framework partially established means the repetition and reconsolidation process can be more effective in in further building and refining the general framework for the task.

There is some important future research that should be conducted on the spacing effect. One important direction for future research is replicating and extending the neuroscience research that relate to memory reconsolidation and the spacing effect. In particular, researchers should examine the neural correlates of consolidation and reconsolidation such as the hippocampal to cortical shift. The reconsolidation account predicts that spaced repetitions will lead to greater hippocampal to cortical shift than massed repetitions. This prediction can be tested through a number of different methods. One method is by examining at what time point a memory becomes independent of the hippocampus. For this test it would be worthwhile to replicate Lehmann and McNamara (2011) as well as testing other tasks and spacing intervals. A second method is by examining the expression of genes associated with neural plasticity, spaced repetitions should be associated with greater expression of genes for neocortical plasticity than massed repetitions. A third method is by using fMRI; on a later test spaced repetitions should lead to stronger connections with neocortical areas and a greater degree of reorganisation than massed repetitions.

In Chapter 1, I noted that adults show a spacing effect for retention of verbal materials but not learning, whereas children seem to show a spacing effect for both learning and retention. Additionally, adults generally show a spacing effect for learning and retaining skills, however, there are some interesting differences in the pattern of spacing which is optimal for enhancing performance. In Chapter 2, I suggested that these observations for learning could be due to an interaction of the reconsolidation process with participant's prior learning. This hypothesis should be tested in future experiments. One possible experiment is adults could learn a maze through spaced repetitions or massed repetitions after

learning 10 different mazes across multiple days (experienced) or not learning 10 mazes (novice). The reconsolidation account predicts that experienced participants will benefit less from spacing for learning than novice participants.

The reconsolidation account of the spacing effect has some potential implications for enhancing learning and memory which should be explored in future research. Some studies of reconsolidation find that providing some new information or change that is part of an already existing memory structure is a potent inducer of memory reconsolidation (Pedreira et al., 2004; Winters et al., 2009). Based on this knowledge, one method for optimising memory might be for students to obtain some initial learning and then allow time for this learning to consolidate and then learn some additional information that elaborates on what was initially learned. For example, in a political history course the traditional method might be to go into depth about the 1908 election and then go into depth about the 1912 election but a method that is optimised based on reconsolidation and the spacing effect would instead give a general overview on the 1908 election and then a general overview of the 1912 election and then go into more depth in the 1908 election and then go into more depth in the 1912 election. This method of spacing and adding novel information into an initial consolidated framework may enhance the memory reconsolidation process, improve long-term retention and potentially increase the speed of learning.

From what we have learned about the spacing effect and memory reconsolidation there are some general principles that we can be mindful of for our learning goals in everyday life. When learning a new skill if possible it is better to spread your learning across multiple days each week. Initially, this might not be as satisfying as you will not see as much progress within a single day but over time you should benefit from greater learning and retention. This might be approached by making the goal of spending 1 hour per day practicing instead of 3 hours on Saturday and Sunday. One should also be mindful of anything that might disrupt the quality of your sleep as this could affect memory consolidation and reconsolidation. Alcohol for example is known to disrupt the natural progression of peoples sleep and this likely influences memory consolidation and learning. For example, suppose you

have the goal of learning a new skill and you plan on practicing every week day and taking the weekend off, until you master the skill you might want to minimise your alcohol consumption on the weekend to avoid slowing down your learning. Finally, one should also keep in mind that you should adjust the intensity of practice or study depending on the nature of a task. Some tasks to make effective progress might require longer more intense periods of concentration. This should be considered when planning out ones learning goals. For example, you might want to learn a new language and learn to play the guitar and you have three hours per day for learning. Since languages can be hard for adults to learn, require attention and memory of many different features (e.g., grammar, vocabulary and pronunciation) and require considerable skill and knowledge to be competent. It might be more effective to first study the language for three hours per day and then once you are competent learn guitar for 3 hours per day, instead of studying the language for an hour and a half per day and learning guitar for an hour and a half.

A broad and interesting question is: does the existence of the spacing effect have a functional advantage and how does this relate to memory reconsolidation? Storing and creating memories has metabolic costs (Burns, Foucaud, & Mery, 2010), additionally an organism does not know whether the information contained within a memory will continue to be useful in the future or if it will no longer be relevant. Furthermore, it is possible that new information may interfere with older more reliable information. The spacing effect may be functionally advantageous because if a particular experience is repeated after a delay that indicates that it is likely to contain useful information for the future (Smid et al., 2007). In terms of consolidation and reconsolidation it may be advantageous to invest a moderate amount of resources in the initial encoding and storage of the memory and then wait and see if any aspect of the experience repeats itself. If some aspect of an experience is repeated more resources are devoted to maintaining the memory but over time if the memory is not repeated it becomes weaker and is forgotten.

Using this perspective on the spacing effect I can examine how it relates to the ITI effect discussed in the Chapter 1. With the ITI effect the time between trials influences how useful the organism perceives the CS is for predicting the US. For the spacing effect the time between repetitions may be influencing how useful the organism perceives the information to be for the future. In this sense both the spacing effect and the ITI effect reflect memory systems seeking to maximise the utility of different types of events.

Conclusion

From my experiments and past research I have a picture of how memory functions in relation to the spacing effect. Over short time scales a delay allows the memory of a stimulus to be partially forgotten and then when a repetition occurs the memory is more difficult to retrieve resulting in more effective processing. Over long timescales a delay allows the memory time for consolidation and when the stimulus is repeated the reconsolidation process rebuilds the memory in a stronger and more durable form. The reconsolidation account of the spacing effect developed in this thesis is a valuable theory as it can explain a number of observations on the spacing effect that previous theories of the spacing effect do not account for. As explained in Chapter 2 the reconsolidation account can explain why spacing under some circumstances enhances learning and under other circumstances does not. The reconsolidation account can explain why learning and retention of specific information is enhanced as well as the generalization of learning. The reconsolidation account explains why when I manipulated the strength of training I observed a similar effect on the spacing effect and the reconsolidation process. Finally, the reconsolidation account can explain why I observed a discontinuity in the influence of retrieval difficulty over different timescales.

References

Aarts, AA, Anderson, JE, Anderson, CJ, Attridge, PR, Attwood, A, & Fedor, Anna. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 1-8.

Abe, Mitsunari, Schambra, Heidi, Wassermann, Eric M, Luckenbaugh, Dave, Schweighofer, Nicolas, & Cohen, Leonardo G. (2011). Reward improves long-term retention of a motor memory through induction of offline memory gains. *Current Biology, 21*(7), 557-562.

Abel, Magdalena, & Bäuml, Karl-Heinz T. (2014). Sleep can reduce proactive interference. *Memory, 22*(4), 332-339.

Aberg, Kristoffer C, Tartaglia, Elisa M, & Herzog, Michael H. (2009). Perceptual learning with Chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep. *Vision research, 49*(16), 2087-2094.

Abrams, Macy L, & Grice, Jon K. (1976). Effects of practice and positional variables in acquisition of a complex psychomotor skill. *Perceptual and motor skills, 43*(1), 203-211.

Alberini, Cristina M. (2005). Mechanisms of memory stabilization: are consolidation and reconsolidation similar or distinct processes? *Trends in neurosciences, 28*(1), 51-56. doi: https://doi.org/10.1016/j.tins.2004.11.001

Alberini, Cristina M. (2011). The role of reconsolidation and the dynamic process of long-term memory formation and storage. *Frontiers in behavioral neuroscience, 5*, 12. doi: 10.3389/fnbeh.2011.00012

Alberini, Cristina M, & LeDoux, Joseph E. (2013). Memory reconsolidation. *Current Biology, 23*(17), R746-R750.

Albert, Ira B, Simmons, James, & Walker, Jerry. (1974). Massed and spaced practice in alpha enhancement. *Perceptual and Motor Skills*.

Alvares, L De Oliveira, Crestani, AP, Cassini, LF, Haubrich, J, Santana, F, & Quillfeldt, JA. (2013). Reactivation enables memory updating, precision-keeping and strengthening: exploring the possible biological roles of reconsolidation. *Neuroscience, 244*, 42-48.

Alvarez, Pablo, & Squire, Larry R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences, 91*(15), 7041-7045.

Ambridge, Ben, Theakston, Anna L, Lieven, Elena VM, & Tomasello, Michael. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development, 21*(2), 174-193.

Andersen, Steven Arild Wuyts, Mikkelsen, Peter Trier, Konge, Lars, Cayé-Thomasen, Per, & Sørensen, Mads Sølvsten. (2016). Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *The Laryngoscope, 126*(2), E74-E79.

Arthur Jr, Winfred, Day, Eric Anthony, Villado, Anton J, Boatman, Paul R, Kowollik, Vanessa, Bennett Jr, Winston, & Bhupatkar, Alok. (2010). The effect of distributed practice on immediate posttraining, and long-term performance on a complex command-and-control simulation task. *Human Performance, 23*(5), 428-445. doi: http://dx.doi.org/10.1080/08959285.2010.515277

Asch, Solomon E. (1969). A reformulation of the problem of associations. *American Psychologist, 24*(2), 92.

Atienza, Mercedes, Cantero, Jose L, & Stickgold, Robert. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience, 16*(1), 53-64.

Baddeley, AD, & Longman, DJA. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21*(8), 627-635.

Bahar, Amir, Dorfman, Nimrod, & Dudai, Yadin. (2004). Amygdalar circuits required for either consolidation or extinction of taste aversion memory are not required for reconsolidation. *European Journal of Neuroscience, 19*(4), 1115-1118.

Bahrick, Harry P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*(3), 296.

Bahrick, Harry P, Bahrick, Lorraine E, Bahrick, Audrey S, & Bahrick, Phyllis E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4*(5), 316-321.

Bahrick, Harry P, & Hall, Lynda K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language, 52*(4), 566-577.

Bahrick, Harry P, & Phelphs, Elizabeth. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(2), 344.

Balota, David A, Duchek, Janet M, & Paullin, Ronda. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and aging, 4*(1), 3.

Begg, Ian, & Green, Craig. (1988). Repetition and trace interaction: Superadditivity. *Memory & cognition, 16*(3), 232-242.

Bekinschtein, Pedro, Cammarota, Martín, Igaz, Lionel Müller, Bevilaqua, Lia RM, Izquierdo, Iván, & Medina, Jorge H. (2007). Persistence of long-term memory storage requires a late protein synthesis-and BDNF-dependent phase in the hippocampus. *Neuron, 53*(2), 261-277.

Bell, Matthew C, Kawadri, Nader, Simone, Patricia M, & Wiseheart, Melody. (2014). Long-term memory, sleep, and the spacing effect. *Memory, 22*(3), 276-283.

Benjamin, Aaron S, & Tullis, Jonathan. (2010). What makes distributed practice effective? *Cognitive psychology, 61*(3), 228-247. doi: https://doi.org/10.1016/j.cogpsych.2010.05.004

Bird, Steve. (2010). Effects of Distributed Practice on the Acquisition of Second Language English Syntax. *Applied Psycholinguistics, 31*(4), 635-650.

Bjork, Robert A, & Allen, Ted W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9*(5), 567-572.

Bjork, Robert A, & Bjork, Elizabeth Ligon. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From learning processes to cognitive processes: Essays in honor of William K. Estes, 2*, 35-67.

Bjork, Robert A, & Whitten, William B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*(2), 173-189.

Bloom, Kristine C, & Shuell, Thomas J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research, 74*(4), 245-248.

Bourne, Lyle E, & Archer, E James. (1956). Time continuously on target as a function of distribution of practice. *Journal of Experimental Psychology, 51*(1), 25.

Bouton, Mark E, & Sunsay, Ceyhun. (2003). Importance of trials versus accumulating time across trials in partially reinforced appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 29*(1), 62.

Bower, Gordon H, Thompson-Schill, Sharon, & Tulving, Endel. (1994). Reducing retroactive interference: An interference analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(1), 51.

Bozon, Bruno, Davis, Sabrina, & Laroche, Serge. (2003). A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval. *Neuron, 40*(4), 695-701.

Brown, Rachel M, & Robertson, Edwin M. (2007). Off-line processing: reciprocal interactions between declarative and procedural memories. *Journal of Neuroscience, 27*(39), 10468-10475.

Burns, James G, Foucaud, Julien, & Mery, Frederic. (2010). Costs of memory: lessons from ‘mini’brains. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20102488.

Cahill, Larry, & McGaugh, James L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in neurosciences, 21*(7), 294-299.

Carpenter, Shana K, & DeLosh, Edward L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & cognition, 34*(2), 268-276.

Censor, Nitzan, Dimyan, Michael A, & Cohen, Leonardo G. (2010). Modification of existing human motor memories is enabled by primary cortical processing during memory reactivation. *Current Biology, 20*(17), 1545-1549.

Cepeda, Nicholas J, Coburn, Noriko, Rohrer, Doug, Wixted, John T, Mozer, Michael C, & Pashler, Harold. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental psychology, 56*(4), 236-246. doi: 10.1027/1618-3169.56.4.236

Cepeda, Nicholas J, Pashler, Harold, Vul, Edward, Wixted, John T, & Rohrer, Doug. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin, 132*(3), 354.

Cepeda, Nicholas J, Vul, Edward, Rohrer, Doug, Wixted, John T, & Pashler, Harold. (2008). Spacing effects in learning a temporal ridgeline of optimal retention. *Psychological science, 19*(11), 1095-1102. doi: https://doi.org/10.1111/j.1467-9280.2008.02209.x

Ceraso, John, & Henderson, Ann. (1965). Unavailability and associative loss in in RI and PI. *Journal of Experimental Psychology, 70*(3), 300.

Childers, Jane B, & Tomasello, Michael. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental psychology*, *38*(6), 967.

Coccoz, V, Maldonado, H, & Delorenzi, A. (2011). The enhancement of reconsolidation with a naturalistic mild stressor improves the expression of a declarative memory in humans. *Neuroscience*, *185*, 61-72.

Commins, Sean, Cunningham, Lorretto, Harvey, Deirdre, & Walsh, Derek. (2003). Massed but not spaced training impairs spatial memory. *Behavioural brain research*, *139*(1), 215-223. doi: https://doi.org/10.1016/S0166-4328(02)00270-X

Cook, Thomas William. (1934). Massed and distributed practice in puzzle solving. *Psychological Review*, *41*(4), 330.

Cook, Thomas William. (1937). Distribution of practice and size of maze pattern. *British Journal of Psychology. General Section*, *27*(3), 303-312.

Cook, Thomas William. (1944). Factors in massed and distributed practice. *Journal of Experimental Psychology*, *34*(4), 325.

Crowder, Robert G. (1976). Principles of learning and memory.

Crystal, Jonathon D, & Babb, Stephanie J. (2008). Spatial memory in rats after 25 hours. *Learning and motivation*, *39*(4), 278-284.

Cull, William L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215-235.

Dail, Teresa K, & Christina, Robert W. (2004). Distribution of practice and metacognition in learning and long-term retention of a discrete motor task. *Research quarterly for exercise and sport*, *75*(2), 148-155. doi: http://dx.doi.org/10.1080/02701367.2004.10609146

Davis, Hasker P, & Squire, Larry R. (1984). Protein synthesis and memory: a review. *Psychological bulletin*, *96*(3), 518.

De Beukelaar, Toon T, Woolley, Daniel G, Alaerts, Kaat, Swinnen, Stephan P, & Wenderoth, Nicole. (2016). Reconsolidation of motor memories is a time-dependent process. *Frontiers in human neuroscience, 10*, 408.

De Beukelaar, Toon T, Woolley, Daniel G, & Wenderoth, Nicole. (2014). Gone for 60 seconds: Reactivation length determines motor memory degradation during reconsolidation. *cortex, 59*, 138-145. doi: https://doi.org/10.1016/j.cortex.2014.07.008

De Win, Gunter, Van Bruwaene, Siska, De Ridder, Dirk, & Miserez, Marc. (2013). The optimal frequency of endoscopic skill labs for training and skill retention on suturing: a randomized controlled trial. *Journal of surgical education, 70*(3), 384-393.

Dębiec, J, & Ledoux, JE. (2004). Disruption of reconsolidation but not consolidation of auditory fear conditioning by noradrenergic blockade in the amygdala. *Neuroscience, 129*(2), 267-272.

Delaney, Peter F, & Verkoeijen, Peter PJL. (2009). Rehearsal strategies can enlarge or diminish the spacing effect: Pure versus mixed lists and encoding strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1148.

Delaney, Peter F, Verkoeijen, Peter PJL, & Spirgel, Arie. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of learning and motivation, 53*, 63-147. doi: https://doi.org/10.1016/S0079-7421(10)53003-2

Demant, Helle, Ladewig, Jan, Balsby, Thorsten JS, & Dabelsteen, Torben. (2011). The effect of frequency and duration of training sessions on acquisition and long-term memory in dogs. *Applied Animal Behaviour Science, 133*(3), 228-234. doi: https://doi.org/10.1016/j.applanim.2011.05.010

Dempster, Frank N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1*(4), 309-330.

Detert, Julia A, Kampa, Nicholas D, & Moyer Jr, James R. (2008). Differential effects of training intertrial interval on acquisition of trace and long-delay fear conditioning in rats. *Behavioral neuroscience, 122*(6), 1318.

Donovan, John J, & Radosevich, David J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*(5), 795.

Drosopoulos, Spyridon, Schulze, Claudia, Fischer, Stefan, & Born, Jan. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General, 136*(2), 169.

Easley, Howard. (1937). The curve of forgetting and the distribution of practice. *Journal of Educational Psychology, 28*(6), 474.

Ebbinghaus, Hermann. (1885/1964). Memory (HA Ruger & CE Bussenius, Trans.). *New York: Teachers College.(Original work published as Das Gedächtnis, 1885)*.

Ericksen, Stanford Clark. (1942). Variability of attack in massed and distributed practice. *Journal of Experimental Psychology, 31*(4), 339.

Estes, William K. (1955). Statistical theory of distributional phenomena in learning. *Psychological review, 62*(5), 369. doi: http://dx.doi.org/10.1037/h0046888

Fischer, Stefan, & Born, Jan. (2009). Anticipated reward enhances offline learning during sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1586.

Fischer, Stefan, Drosopoulos, Spyridon, Tsen, Jim, & Born, Jan. (2006). Implicit learning–explicit knowing: a role for sleep in memory system interaction. *Journal of Cognitive Neuroscience, 18*(3), 311-319.

Forcato, Cecilia , Argibay, Pablo F, Pedreira, María E, & Maldonado, Hector. (2009). Human reconsolidation does not always occur when a memory is retrieved: the relevance of the reminder structure. *Neurobiology of Learning and Memory, 91*(1), 50-57.

Forcato, Cecilia, Burgos, Valeria L, Argibay, Pablo F, Molina, Victor A, Pedreira, María E, & Maldonado, Hector. (2007). Reconsolidation of declarative memory in humans. *Learning & Memory, 14*(4), 295-303.

Frankland, Paul W, & Bontempi, Bruno. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience, 6*(2), 119.

Franklin, Joseph C, & Brozek, Josef. (1947). The relation between distribution of practice and learning efficiency in psychomotor performance. *Journal of experimental psychology, 37*(1), 16.

Friedrich, Manuela, Wilhelm, Ines, Born, Jan, & Friederici, Angela D. (2015). Generalization of word meanings during infant sleep. *Nature communications, 6*.

Gaab, Nadine, Paetzold, Miriam, Becker, Markus, Walker, Matthew P, & Schlaug, Gottfried. (2004). The influence of sleep on auditory learning: a behavioral study. *Neuroreport, 15*(4), 731-734.

Gais, Steffen, Mölle, Matthias, Helms, Kay, & Born, Jan. (2002). Learning-dependent increases in sleep spindle density. *Journal of Neuroscience, 22*(15), 6830-6834.

Gais, Steffen, Plihal, Werner, Wagner, Ullrich, & Born, Jan. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature neuroscience, 3*(12), 1335-1339.

Gallagher, Anthony G, Jordan-Black, Julie Anne, & O'Sullivan, Gerald C. (2012). Prospective, randomized assessment of the acquisition, maintenance, and loss of laparoscopic skills. *Annals of surgery, 256*(2), 387-393.

Galluccio, Lissa, & Rovee-Collier, Carolyn. (2006). Nonuniform effects of reinstatement within the time window. *Learning and Motivation, 37*(1), 1-17.

Garcia, JA, Moreno, FJ, Reina, R, Menayo, R, & Fuentes, JP. (2008). Analysis of effects of distribution of practice in learning and retention of a continuous and a discrete skill presented on a computer. *Perceptual and motor skills, 107*(1), 261-272.

Garrett, HENRY E. (1940). Variability in learning under massed and spaced practice. *Journal of Experimental Psychology, 26*(6), 547.

Gerbier, Emilie, Toppino, Thomas C, & Koenig, Olivier. (2015). Optimising retention through multiple study opportunities over days: The benefit of an expanding schedule of repetitions. *Memory, 23*(6), 943-954.

Gershman, Samuel J, Schapiro, Anna C, Hupbach, Almut, & Norman, Kenneth A. (2013). Neural context reinstatement predicts memory misattribution. *Journal of Neuroscience, 33*(20), 8590-8595.

Glenberg, Arthur M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15*(1), 1-16.

Glenberg, Arthur M. (1977). Influences of retrieval processes on the spacing effect in free recall. *Journal of Experimental Psychology: Human Learning and Memory, 3*(3), 282.

Glenberg, Arthur M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*(2), 95-112.

Glenberg, Arthur M, & Lehmann, Thomas S. (1980). Spacing repetitions over 1 week. *Memory & Cognition, 8*(6), 528-538.

Glenberg, Arthur M, & Smith, Steven M. (1981). Spacing repetitions and solving problems are not the same. *Journal of Verbal Learning and Verbal Behavior, 20*(1), 110-119.

Glover, John A. (1989). The" testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392.

Gluckman, Maxie, Vlach, Haley A, & Sandhofer, Catherine M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology, 28*(2), 266-273.

Goedert, Kelly M, & Miller, Jason. (2008). Spacing practice sessions across days earlier rather than later in training improves performance of a visuomotor skill. *Experimental brain research, 189*(2), 189-197.

Goelet, Philip, Castellucci, Vincent F, Schacher, Samuel, & Kandel, Eric R. (1986). The long and the short of long-term memory: a molecular framework. *Nature*.

Goeleven, Ellen, De Raedt, Rudi, Leyman, Lemke, & Verschuere, Bruno. (2008). The Karolinska directed emotional faces: a validation study. *Cognition and emotion, 22*(6), 1094-1118.

Golkar, Armita, Bellander, Martin, Olsson, Andreas, & Öhman, Arne. (2012). Are fear memories erasable?–reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli. *Frontiers in behavioral neuroscience, 6*, 80.

Gómez, Rebecca L, Bootzin, Richard R, & Nadel, Lynn. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17*(8), 670-674.

Goodrick, Charles L. (1973). Maze learning of mature-young and aged rats as a function of distribution of practice. *Journal of Experimental Psychology, 98*(2), 344.

Goossens, Nicole AMC, Camp, Gino, Verkoeijen, Peter PJL, Tabbers, Huib K, & Zwaan, Rolf A. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology, 24*(8), 965-971.

Greene, Robert L. (2014). *Human memory: Paradigms and paradoxes*: Psychology Press.

Hagman, Joseph D. (1980). Effects of training schedule and equipment variety on retention and transfer of maintenance skill: DTIC Document.

Hamann, Stephan B, Ely, Timothy D, Grafton, Scott T, & Kilts, Clinton D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature neuroscience, 2*(3), 289-293.

Hardwicke, Tom E, Taqi, Mahdi, & Shanks, David R. (2016). Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proceedings of the National Academy of Sciences, 113*(19), 5206-5211.

Harmon, John M, & Miller, Arthur G. (1950). Time patterns in motor learning. *Research Quarterly. American Association for Health, Physical Education and Recreation, 21*(3), 182-187.

Hartshorn, Kristin, Rovee-Collier, Carolyn, Gerhardstein, Peter, Bhatt, Ramesh S, Wondoloski, Teresa L, Klein, Pamela, . . . Campos-de-Carvalho, Mara. (1998). The ontogeny of long-term memory over the first year-and-a-half of life. *Developmental Psychobiology, 32*(2), 69-89.

Hartshorn, Kristin, Wilk, Amy E, Muller, Katherine L, & Rovee-Collier, Carolyn. (1998). An expanding training series protracts retention for 3-month-old infants. *Developmental psychobiology, 33*(3), 271-282.

Heidt, Constance Tamara, Arbuthnott, Katherine D, & Price, Heather L. (2016). The effects of distributed learning on enhanced cognitive interview training. *Psychiatry, Psychology and Law, 23*(1), 47-61.

Hintzman, Douglas L. (1974). Theoretical implications of the spacing effect.

Hintzman, Douglas L, Block, Richard A, & Summers, Jeffery J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior, 12*(2), 229-238.

Hintzman, Douglas L, & Rogers, Miriam K. (1973). Spacing effects in picture memory. *Memory & Cognition, 1*(4), 430-434.

Hintzman, Douglas L, Summers, Jeffrey J, & Block, Richard A. (1975). Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory, 1*(1), 31.

Hogan, Robert M, & Kintsch, Walter. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10*(5), 562-567.

Hovland, Carl I. (1938). Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice. *Journal of Experimental Psychology, 22*(3), 201.

Hovland, Carl I. (1940). Experimental studies in rote-learning theory. VI. Comparison of retention following learning to same criterion by massed and distributed practice. *Journal of Experimental Psychology, 26*(6), 568.

Hsu, Vivian C. (2010). Time windows in retention over the first year-and-a-half of life: Spacing effects. *Developmental psychobiology, 52*(8), 764-774.

Hudson, Judith A, & Sheffield, Ellyn G. (1998). Déjà vu all over again: Effects of reenactment on toddlers' event memory. *Child Development, 69*(1), 51-67.

Hupbach, Almut, Gomez, Rebecca, Hardt, Oliver, & Nadel, Lynn. (2007). Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learning & Memory, 14*(1-2), 47-53.

Inda, Maria Carmen, Muravieva, Elizaveta V, & Alberini, Cristina M. (2011). Memory retrieval and the passage of time: from reconsolidation and strengthening to extinction. *The Journal of Neuroscience, 31*(5), 1635-1643.

Janiszewski, Chris, Noel, Hayden, & Sawyer, Alan G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of consumer research, 30*(1), 138-149.

Ji, Daoyun, & Wilson, Matthew A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience, 10*(1), 100-107.

Kahana, Michael J, & Howard, Marc W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review, 12*(1), 159-164.

Kang, Sung Gu, Ryu, Byung Ju, Yang, Kyung Sook, Ko, Young Hwii, Cho, Seok, Kang, Seok Ho, . . . Cheon, Jun. (2015). An effective repetitive training schedule to achieve skill proficiency using a novel robotic virtual reality simulator. *Journal of surgical education, 72*(3), 369-376.

Karni, Avi, Meyer, Gundela, Rey-Hipolito, Christine, Jezzard, Peter, Adams, Michelle M, Turner, Robert, & Ungerleider, Leslie G. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences, 95*(3), 861-868.

Karni, Avi, Tanne, David, Rubenstein, Barton S, Askenasy, Jean JM, & Sagi, Dov. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science, 265*(5172), 679-683.

Kauffeld, Simone, & Lehmann-Willenbrock, Nale. (2010). Sales training: effects of spaced practice on training transfer. *Journal of European industrial training, 34*(1), 23-37.

Keppel, Geoffrey. (1964). Facilitation in short-and long-term retention of paired associates following distributed practice in learning. *Journal of Verbal Learning and Verbal Behavior, 3*(2), 91-111.

Keppel, Geoffrey. (1967). A reconsideration of the extinction-recovery theory. *Journal of Verbal Learning and Verbal Behavior, 6*(4), 476-486.

Kerfoot, B Price, Fu, Yineng, Baker, Harley, Connelly, Donna, Ritchey, Michael L, & Genega, Elizabeth M. (2010). Online spaced education generates transfer and improves long-term retention of diagnostic skills: a randomized controlled trial. *Journal of the American College of Surgeons, 211*(3), 331-337. e331.

Kılıç, Aslı, Hoyer, William J, & Howard, Marc W. (2013). Effects of spacing of item repetitions in continuous recognition memory: Does item retrieval difficulty promote item retention in older

adults? *Experimental aging research, 39*(3), 322-341. doi: http://dx.doi.org/10.1080/0361073X.2013.779200

Kindt, Merel, & Soeter, Marieke. (2013). Reconsolidation in a human fear conditioning study: a test of extinction as updating mechanism. *Biological psychology, 92*(1), 43-50.

Knapp, Clyde G, & Dixon, W Robert. (1950). Learning to juggle: I. A study to determine the effect of two different distributions of practice on learning efficiency. *Research Quarterly. American Association for Health, Physical Education and Recreation, 21*(3), 331-340.

Knapp, Clyde G, Dixon, W Robert, & Lazier, Murney. (1958). Learning to juggle: III. A study of performance by two different age groups. *Research Quarterly. American Association for Health, Physical Education and Recreation, 29*(1), 32-36.

Kraemer, Philipp J, & Randall, Christopher K. (1995). Spatial learning in preweanling rats trained in a Morris water maze. *Psychobiology, 23*(2), 144-152.

Küpper-Tetzel, Carolina E, & Erdfelder, Edgar. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory, 20*(1), 37-47.

Küpper-Tetzel, Carolina E, Erdfelder, Edgar, & Dickhäuser, Oliver. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science, 42*(3), 373-388.

Küpper-Tetzel, Carolina E, Kapler, Irina V, & Wiseheart, Melody. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & cognition, 42*(5), 729-741.

Kuriyama, Kenichi, Stickgold, Robert, & Walker, Matthew P. (2004). Sleep-dependent learning and motor-skill complexity. *Learning & Memory, 11*(6), 705-713.

Kusunose, Ryo, & Yamanobe, Akira. (2002). The effect of training schedule on learned tasks in yearling horses. *Applied Animal Behaviour Science, 78*(2), 225-233.

Lahl, Olaf, Wispel, Christiane, Willigens, Bernadette, & Pietrowsky, Reinhard. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of sleep research, 17*(1), 3-10.

Landauer, Thomas K. (1969). Reinforcement as consolidation. *Psychological Review, 76*(1), 82.

Lashley, KS. (1915). The acquisition of skill in archery. *Department of Marine Biology of the Carnegie Institution of Washington, 7*, 105-128.

Laureys, Steven, Peigneux, Philippe, Phillips, Christophe, Fuchs, Sonia, Degueldre, Christian, Aerts, Joël, . . . Van der Linden, Martial. (2001). Experience-dependent changes in cerebral functional connectivity during human rapid eye movement sleep. *Neuroscience, 105*(3), 521-525.

Lee, Jonathan LC. (2008). Memory reconsolidation mediates the strengthening of memories by additional learning. *Nature neuroscience, 11*(11), 1264-1266. doi: http://dx.doi.org/10.1038/nn.2205

Lee, Timothy D, & Genovese, Elizabeth D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport, 59*(4), 277-287.

Lehmann, Hugo, & McNamara, Kathryn C. (2011). Repeatedly reactivated memories become more resistant to hippocampal damage. *Learning & Memory, 18*(3), 132-135.

Lehmann, Hugo, Sparks, Fraser T, Spanswick, Simon C, Hadikin, Crystal, McDonald, Robert J, & Sutherland, Robert J. (2009). Making context memories independent of the hippocampus. *Learning & Memory, 16*(7), 417-420.

Lundqvist, Daniel, Flykt, Anders, & Öhman, Arne. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*(1998).

Maddox, Geoffrey B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: a case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology*, 1-23.

Maquet, Pierre, Laureys, Steven, Peigneux, Philippe, Fuchs, Sonia, Petiau, Christophe, Phillips, Christophe, . . . Meulemans, Thierry. (2000). Experience-dependent changes in cerebral activation during human REM sleep. *Nature neuroscience*, *3*(8), 831-836.

Marshall, Lisa, Helgadóttir, Halla, Mölle, Matthias, & Born, Jan. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature*, *444*(7119), 610-613.

Massey, M Dorothy. (1959). The significance of interpolated time intervals on motor learning. *Research Quarterly. American Association for Health, Physical Education and Recreation*, *30*(2), 189-201.

Mazza, Stéphanie, Gerbier, Emilie, Gustin, Marie-Paule, Kasikci, Zumrut, Koenig, Olivier, Toppino, Thomas C, & Magnin, Michel. (2016). Relearn faster and retain longer: Along with practice, sleep makes perfect. *Psychological science*, *27*(10), 1321-1330.

McGaugh, James L. (2000). Memory--a century of consolidation. *Science*, *287*(5451), 248-251.

McGaugh, James L, & Cole, James M. (1965). Age and strain differences in the effect of distribution of practice on maze learning. *Psychonomic Science*, *2*(1-12), 253-254.

McGeoch, John A, & Irion, Arthur L. (1952). The psychology of human learning.

McNaughton, BL, Douglas, RM, & Goddard, GV. (1991). Dependence on REM sleep of overnight improvement of a perceptual skill. *Network*, *2*, 371.

Mednick, Sara C, Cai, Denise J, Shuman, Tristan, Anagnostaras, Stephan, & Wixted, John T. (2011). An opportunistic theory of cellular and systems consolidation. *Trends in neurosciences*, *34*(10), 504-514.

Mednick, Sara, Nakayama, Ken, & Stickgold, Robert. (2003). Sleep-dependent learning: a nap is as good as a night. *Nature neuroscience, 6*(7), 697-698. doi: http://dx.doi.org/10.1038/nn1078

Meiri, Noam, & Rosenblum, Kobi. (1998). Lateral ventricle injection of the protein synthesis inhibitor anisomycin impairs long-term memory in a spatial memory task. *Brain research, 789*(1), 48-55.

Melton, Arthur W. (1967). Repetition and retrieval from memory. *Science, 158*(3800), 532-532.

Menzel, Randolf, Manz, Gisela, Menzel, Rebecca, & Greggers, Uwe. (2001). Massed and spaced learning in honeybees: the role of CS, US, the intertrial interval, and the test interval. *Learning & Memory, 8*(4), 198-208.

Metalis, Spero A. (1985). Effects of massed versus distributed practice on acquisition of video game skill. *Perceptual and Motor Skills, 61*(2), 457-458.

Meyer, Iben, & Ladewig, Jan. (2008). The relationship between number of training sessions per week and learning in dogs. *Applied animal behaviour science, 111*(3), 311-320.

Miles, Scott W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System, 42*, 412-428.

Miller, Ralph R, & Marlin, Nancy A. (2014). Amnesia following electroconvulsive shock. *Functional disorders of memory*, 143-178.

Milner, Peter M. (1989). A cell assembly theory of hippocampal amnesia. *Neuropsychologia, 27*(1), 23-30.

Mitchell, Erica L, Lee, Dae Y, Sevdalis, Nick, Partsafas, Aaron W, Landry, Gregory J, Liem, Timothy K, & Moneta, Gregory L. (2011). Evaluation of distributed practice schedules on retention of a newly acquired surgical skill: a randomized trial. *The American Journal of Surgery, 201*(1), 31-39.

Moinzadeh, Ahmad Reza, Talebinezhad, Mohammad Reza, & Behazin, Arash. (2008). Exposure density in relation to learning and retention in EFL. *The International Journal of Humanities, 15*(2), 71-98.

Molloy, Katharine, Moore, David R, Sohoglu, Ediz, & Amitay, Sygal. (2012). Less is more: latent learning is maximized by shorter training sessions in auditory perceptual learning. *PloS one, 7*(5), e36929.

Morris, RGM, & Doyle, J. (1985). Successive incompatible tasks: evidence for separate subsystems for storage of spatial knowledge. *Electrical activity of the archicortex (ed. Buzsaki G, Vanderwolf CH)*, 281-293.

Morris, Richard GM, Inglis, Jennifer, Ainge, James A, Olverman, Henry J, Tulloch, Jane, Dudai, Yadin, & Kelly, Paul AT. (2006). Memory reconsolidation: sensitivity of spatial memory to inhibition of protein synthesis in dorsal hippocampus during encoding and retrieval. *Neuron, 50*(3), 479-489.

Moss, Vanessa Dawn. (1996). *The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student.* ProQuest Information & Learning.

Moulton, Carol-Anne E, Dubrowski, Adam, MacRae, Helen, Graham, Brent, Grober, Ethan, & Reznick, Richard. (2006). Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Annals of surgery, 244*(3), 400.

Murphree, Thomas Rutherford. (1971). Effects of massed and distributed practice upon motor learning and retention of a novel gross motor task.

Nader, Karim, & Hardt, Oliver. (2009). A single standard for memory: the case for reconsolidation. *Nature Reviews Neuroscience, 10*(3), 224-234.

Nader, Karim, Schafe, Glenn E, & Le Doux, Joseph E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature, 406*(6797), 722-726. doi: http://dx.doi.org/10.1038/35021052

Nelson, Thomas O, & Dunlosky, John. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2*(3), 325-335.

Nielson, Harold C. (1968). Evidence that electroconvulsive shock alters memory retrieval rather than memory consolidation. *Experimental Neurology, 20*(1), 3-20.

Oudiette, Delphine, & Paller, Ken A. (2013). Upgrading the sleeping brain with targeted memory reactivation. *Trends in cognitive sciences, 17*(3), 142-149.

Paik, Jaehyon, & Ritter, Frank E. (2015). Evaluating a range of learning schedules: hybrid training schedules may be as good as or better than distributed practice for some tasks. *Ergonomics*, 1-15.

Paivio, Allan. (1974). Spacing of repetitions in the incidental and intentional free recall of pictures and words. *Journal of Verbal Learning and Verbal Behavior, 13*(5), 497-511.

Pashler, Harold, Cepeda, Nicholas, Lindsey, Robert V, Vul, Ed, & Mozer, Michael C. (2009). *Predicting the optimal spacing of study: A multiscale context model of memory.* Paper presented at the Advances in neural information processing systems.

Pavlides, Constantine, & Winson, Jonathan. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *J Neurosci, 9*(8), 2907-2918.

Payne, Jessica D, Chambers, Alexis M, & Kensinger, Elizabeth A. (2012). Sleep promotes lasting changes in selective memory for emotional scenes. *Frontiers in Integrative Neuroscience, 6*.

Pechstein, Louis Augustus. (1917). Whole vs. part methods in motor learning. A comparative story. *The Psychological Monographs, 23*(2), i.

Pedreira, María Eugenia, Pérez-Cuesta, Luis María, & Maldonado, Héctor. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learning & Memory, 11*(5), 579-585.

Peterson, Lloyd R, Wampler, Richard, Kirkpatrick, Meredith, & Saltzman, Dorothy. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology, 66*(2), 206.

Poe, Gina R, Nitz, Douglas A, McNaughton, Bruce L, & Barnes, Carol A. (2000). Experience-dependent phase-reversal of hippocampal neuron firing during REM sleep. *Brain research, 855*(1), 176-180.

Postman, Leo, & Phillips, Laura W. (1954). Studies in incidental learning: I. The effects of crowding and isolation. *Journal of experimental psychology, 48*(1), 48.

Postman, Leo, Stark, Karen, & Fraser, Janat. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior, 7*(3), 672-694.

Pyc, Mary A, & Rawson, Katherine A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437-447. doi: https://doi.org/10.1016/j.jml.2009.01.004

Raaijmakers, Jeroen GW. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27*(3), 431-452. doi: https://doi.org/10.1016/S0364-0213(03)00007-7

Rawson, Katherine A. (2012). Why do rereading lag effects depend on test delay? *Journal of Memory and Language, 66*(4), 870-884.

Rawson, Katherine A, & Kintsch, Walter. (2005). Rereading Effects Depend on Time of Test. *Journal of Educational Psychology, 97*(1), 70.

Rea, Cornelius P, & Modigliani, Vito. (1987). The spacing effect in 4-to 9-year-old children. *Memory & Cognition, 15*(5), 436-443.

Renshaw, Samuel, & Schwarzbek, William C. (1938a). The dependence of the form of the pursuitmeter learning function on the length of the inter-practice rests: I. Experimental. *The Journal of General Psychology, 18*(1), 3-16.

Renshaw, Samuel, & Schwarzbek, William C. (1938b). The dependence of the form of the pursuitmeter learning function on the length of the inter-practice rests: II. Theoretical. *The Journal of General Psychology, 18*(1), 17-29.

Rick, John Th, Murphy, M Paul, Ivy, Gwen O, & Milgram, Norton W. (1996). Short intertrial intervals impair water maze performance in old Fischer 344 rats. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 51*(4), B253-B260.

Roediger, III, Henry L. (2008). Relativity of remembering: Why the laws of memory vanished. *Annu. Rev. Psychol., 59*, 225-254.

Roediger III, Henry L, & Karpicke, Jeffrey D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science, 17*(3), 249-255.

Romkema, Sietske, Bongers, Raoul M, & van der Sluis, Corry K. (2015). Influence of Inter-Training Intervals on Intermanual Transfer Effects in Upper-Limb Prosthesis Training: A Randomized Pre-Posttest Study. *PloS one, 10*(6), e0128747.

Rossato, Janine I, Bevilaqua, Lia RM, Myskiw, Jociane C, Medina, Jorge H, Izquierdo, Iván, & Cammarota, Martín. (2007). On the role of hippocampal protein synthesis in the consolidation and reconsolidation of object recognition memory. *Learning & Memory, 14*(1-2), 36-46.

Rovee-Collier, Carolyn. (1995). Time windows in cognitive development. *Developmental Psychology, 31*(2), 147.

Rovee-Collier, Carolyn, Evancio, Sharon, & Earley, Linda A. (1995). The time window hypothesis: Spacing effects. *Infant Behavior and Development, 18*(1), 69-78.

Rubin-Rabson, Grace. (1940). Studies in the psychology of memorizing piano music: II. A comparison of massed and distributed practice. *Journal of Educational Psychology, 31*(4), 270.

Rubin, L, Oppegard, C, & Hintz, HF. (1980). The effect of varying the temporal distribution of conditioning trials on equine learning behavior. *Journal of animal science, 50*(6), 1184-1187. doi: 10.2527/jas1980.5061184x

Ruch, Theodore C. (1928). Factors influencing the relative economy of massed and distributed practice in learning. *Psychological Review, 35*(1), 19.

Sara, Susan J. (2000). Retrieval and reconsolidation: toward a neurobiology of remembering. *Learning & Memory, 7*(2), 73-84.

Sara, Susan J. (2010). Reactivation, retrieval, replay and reconsolidation in and out of sleep: connecting the dots. *Frontiers in behavioral neuroscience, 4*, 185.

Savion-Lemieux, Tal, & Penhune, Virginia B. (2005). The effects of practice and delay on motor skill learning and retention. *Experimental Brain Research, 161*(4), 423-431.

Schafe, Glenn E, & LeDoux, Joseph E. (2000). Memory consolidation of auditory pavlovian fear conditioning requires protein synthesis and protein kinase A in the amygdala. *J Neurosci, 20*(18), RC96.

Schafe, Glenn E, Nadel, Nicole V, Sullivan, Gregory M, Harris, Alexander, & LeDoux, Joseph E. (1999). Memory consolidation for contextual and auditory fear conditioning is dependent on protein synthesis, PKA, and MAP kinase. *Learning & Memory, 6*(2), 97-110.

Schendel, Joel D, & Hagman, Joseph D. (1982). On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology, 67*(5), 605.

Schwabe, Lars, & Wolf, Oliver T. (2009). New episodic learning interferes with the reconsolidation of autobiographical memories. *PLoS One, 4*(10), e7519.

Schwartz, Richard G, & Terrell, Brenda Y. (1983). The role of input frequency in lexical acquisition. *Journal of child language, 10*(01), 57-64.

Sederberg, Per B, Gershman, Samuel J, Polyn, Sean M, & Norman, Kenneth A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic bulletin & review, 18*(3), 455-468.

Shaughnessy, John J. (1976). Persistence of the spacing effect in free recall under varying incidental learning conditions. *Memory & Cognition, 4*(4), 369-377.

Shaughnessy, John J, Zimmerman, Joel, & Underwood, Benton J. (1972). Further evidence on the MP-DP effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 11*(1), 1-12.

Shea, Charles H, Lai, Qin, Black, Charles, & Park, Jin-Hoon. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human movement science, 19*(5), 737-760.

Shebilske, Wayne L, Goettl, Barry P, Corrington, Kip, & Day, Eric Anthony. (1999). Interlesson spacing and task-related processing during complex skill acquisition. *Journal of Experimental Psychology: Applied, 5*(4), 413.

Simmons, Amy L. (2011). Distributed practice and procedural memory consolidation in musicians' skill learning. *Journal of Research in Music Education*, 0022429411424798.

Simone, Patricia M, Bell, Matthew C, & Cepeda, Nicholas J. (2013). Diminished but not forgotten: Effects of aging on magnitude of spacing effect benefits. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 68*(5), 674-680.

Sissons, Heather T, & Miller, Ralph R. (2009). Overexpectation and trial massing. *Journal of Experimental Psychology: Animal Behavior Processes, 35*(2), 186.

Sisti, Helene M, Glass, Arnold L, & Shors, Tracey J. (2007). Neurogenesis and the spacing effect: learning over time enhances memory and the survival of new neurons. *Learning & memory, 14*(5), 368-375.

Smid, Hans M, Wang, Guohong, Bukovinszky, Tibor, Steidle, Johannes LM, Bleeker, Maartje AK, van Loon, Joop JA, & Vet, Louise EM. (2007). Species-specific acquisition and consolidation of long-term memory in parasitic wasps. *Proceedings of the Royal Society of London B: Biological Sciences, 274*(1617), 1539-1546.

Sobel, Hailey S, Cepeda, Nicholas J, & Kapler, Irina V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25*(5), 763-767. doi: 10.1002/acp.1747

Soeter, Marieke, & Kindt, Merel. (2011). Disrupting reconsolidation: pharmacological and behavioral manipulations. *Learning & Memory, 18*(6), 357-366.

Spreng, Matthieu, Rossier, Jérôme, & Schenk, Françoise. (2002). Spaced training facilitates long-term retention of place navigation in adult but not in adolescent rats. *Behavioural brain research, 128*(1), 103-108.

Spruit, Edward N, Band, Guido PH, & Hamming, Jaap F. (2014). Increasing efficiency of surgical training: effects of spacing practice on skill acquisition and retention in laparoscopy training. *Surgical endoscopy*, 1-9.

Stafford, Tom, & Dewar, Michael. (2014). Tracing the trajectory of skill learning with a very large sample of online game players. *Psychological science, 25*(2), 511-518.

Stelmach, George E. (1969). Efficiency of motor learning as a function of intertrial rest. *Research Quarterly. American Association for Health, Physical Education and Recreation, 40*(1), 198-202.

Stickgold, Robert. (2006). Neuroscience: a memory boost while you sleep. *Nature, 444*(7119), 559-560.

Stickgold, Robert, & Walker, Matthew P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep medicine, 8*(4), 331-343.

Stickgold, Robert, & Walker, Matthew P. (2013). Sleep-dependent memory triage: evolving generalization through selective processing. *Nature neuroscience, 16*(2), 139-145.

Strange, Bryan A, Hurlemann, R, & Dolan, Raymond J. (2003). An emotion-induced retrograde amnesia in humans is amygdala-and β-adrenergic-dependent. *Proceedings of the National Academy of Sciences, 100*(23), 13626-13631.

Strange, Bryan A, Kroes, Marijn CW, Fan, Judith E, & Dolan, Raymond J. (2010). Emotion causes targeted forgetting of established memories. *Frontiers in behavioral neuroscience, 4*.

Sugawara, Sho K, Tanaka, Satoshi, Okazaki, Shuntaro, Watanabe, Katsumi, & Sadato, Norihiro. (2012). Social rewards enhance offline improvements in motor skill. *PLoS One, 7*(11), e48174.

Sunsay, Ceyhun, & Bouton, Mark E. (2008). Analysis of a trial-spacing effect with relatively long intertrial intervals. *Learning & Behavior, 36*(2), 104-115.

Suzuki, Akinobu, Josselyn, Sheena A, Frankland, Paul W, Masushige, Shoichi, Silva, Alcino J, & Kida, Satoshi. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *The Journal of neuroscience, 24*(20), 4787-4795. doi: https://doi.org/10.1523/JNEUROSCI.5491-03.2004

Suzuki, Yuichi, & DeKeyser, Robert. (2015). Effects of distributed practice on the proceduralization of morphology. *Language Teaching Research*, 1362168815617334.

Taubenfeld, Stephen M, Milekic, Maria H, Monti, Barbara, & Alberini, Cristina M. (2001). The consolidation of new but not reactivated memory requires hippocampal C/EBPβ. *Nature neuroscience, 4*(8), 813-818.

Thios, Samuel J, & D'Agostino, Paul R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15*(5), 529-536.

Toppino, Thomas C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition, 19*(2), 159-167.

Toppino, Thomas C. (1993). The spacing effect in preschool children's free recall of pictures and words. *Bulletin of the Psychonomic Society, 31*(1), 27-30.

Toppino, Thomas C, & Bloom, Lance C. (2002). The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 437.

Toppino, Thomas C, Kasserman, Jane E, & Mracek, Wayne A. (1991). The effect of spacing repetitions on the recognition memory of young children and adults. *Journal of experimental child psychology, 51*(1), 123-138.

Tse, Dorothy, Langston, Rosamund F, Kakeyama, Masaki, Bethus, Ingrid, Spooner, Patrick A, Wood, Emma R, . . . Morris, Richard GM. (2007). Schemas and memory consolidation. *Science, 316*(5821), 76-82. doi: 10.1126/science.1135935

Tse, Dorothy, Takeuchi, Tomonori, Kakeyama, Masaki, Kajii, Yasushi, Okuno, Hiroyuki, Tohyama, Chiharu, . . . Morris, Richard GM. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science, 333*(6044), 891-895. doi: 10.1126/science.1205274

Tulving, Endel, & Watkins, Michael J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior, 13*(2), 181-193.

Vander Linde, Eleanor, Morrongiello, Barbara A, & Rovee-Collier, Carolyn. (1985). Determinants of retention in 8-week-old infants. *Developmental Psychology, 21*(4), 601.

Verdaasdonk, EGG, Stassen, LPS, Van Wijk, RPJ, & Dankelman, J. (2007). The influence of different training schedules on the learning of psychomotor skills for endoscopic surgery. *Surgical endoscopy, 21*(2), 214-219.

Verkoeijen, Peter PJL, Rikers, Remy MJP, & Özsoy, Binnur. (2008). Distributed rereading can hurt the spacing effect in text memory. *Applied Cognitive Psychology, 22*(5), 685-695. doi: 10.1002/acp.1388

Verkoeijen, Peter PJL, Rikers, Remy MJP, & Schmidt, Henk G. (2005). Limitations to the spacing effect: Demonstration of an inverted u-shaped relationship between interrepetition spacing and free recall. *Experimental Psychology, 52*(4), 257-263.

Vilberg, Kaia L, & Davachi, Lila. (2013). Perirhinal-hippocampal connectivity during reactivation is a marker for object-based memory consolidation. *Neuron, 79*(6), 1232-1242.

Vlach, Haley A, & Sandhofer, Catherine M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child development, 83*(4), 1137-1144.

Wahlheim, Christopher N, Maddox, Geoffrey B, & Jacoby, Larry L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: Sufficient but not necessary. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(1), 94.

Walker, Matthew P, Brakefield, Tiffany, Hobson, J Allan, & Stickgold, Robert. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425*(6958), 616-620.

Walker, Matthew P, & Stickgold, Robert. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44*(1), 121-133.

Walker, Matthew P, & Stickgold, Robert. (2016). Understanding the boundary conditions of memory reconsolidation. *Proceedings of the National Academy of Sciences of The United States of America, 113*(28), E3991.

Wang, Szu-Han, de Oliveira Alvares, Lucas, & Nader, Karim. (2009). Cellular and systems mechanisms of memory strength as a constraint on auditory fear reconsolidation. *Nature neuroscience, 12*(7), 905-912. doi: 10.1038/nn.2350

Wang, Zuowei, Zhou, Renlai, & Shah, Priti. (2014). Spaced cognitive training promotes training transfer. *Frontiers in human neuroscience, 8*.

Ward, Ryan D, Gallistel, CR, & Balsam, Peter D. (2013). It's the information! *Behavioural processes, 95*, 3-7.

Warriner, Amy Beth, Kuperman, Victor, & Brysbaert, Marc. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods, 45*(4), 1191-1207.

Wichert, Sonja, Wolf, Oliver T, & Schwabe, Lars. (2011). Reactivation, interference, and reconsolidation: Are recent and remote memories likewise susceptible? *Behavioral Neuroscience, 125*(5), 699.

Wickelgren, Wayne A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9*(4), 418-455.

Wilhelm, Ines, Diekelmann, Susanne, Molzow, Ina, Ayoub, Amr, Mölle, Matthias, & Born, Jan. (2011). Sleep selectively enhances memory expected to be of future relevance. *Journal of Neuroscience, 31*(5), 1563-1569.

Wilson, Matthew A, & McNaughton, Bruce L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science, 265*(5172), 676-679.

Winters, Boyer D, Tucci, Mark C, & DaCosta-Furtado, Melynda. (2009). Older and stronger object memories are selectively destabilized by reactivation in the presence of new information. *Learning & Memory, 16*(9), 545-553.

Wixted, John T, & Ebbesen, Ebbe B. (1991). On the form of forgetting. *Psychological science, 2*(6), 409-415.

Wright, Beverly A, & Sabin, Andrew T. (2007). Perceptual learning: how much daily training is enough? *Experimental Brain Research, 180*(4), 727-736.

Xue, Gui, Mei, Leilei, Chen, Chuansheng, Lu, Zhong-Lin, Poldrack, Russell, & Dong, Qi. (2011). Spaced learning enhances subsequent recognition memory by reducing neural repetition suppression. *Journal of cognitive neuroscience, 23*(7), 1624-1633. doi: 10.1162/jocn.2010.21532

Yaroush, Rita, Sullivan, Michael J, & Ekstrand, Bruce R. (1971). Effect of sleep on memory: II. Differential effect of the first and second half of the night. *Journal of experimental psychology, 88*(3), 361.

Yin, Hua, Barnet, Robert C, & Miller, Ralph R. (1994). Trial spacing and trial distribution effects in Pavlovian conditioning: contributions of a comparator mechanism. *Journal of Experimental Psychology: Animal Behavior Processes, 20*(2), 123.

Young, Joseph L. (1971). Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology, 8*(1), 58-81.

Young, Olive G. (1954). Rate of learning in relation to spacing of practice periods in archery and badminton. *Research Quarterly. American Association for Health, Physical Education and Recreation, 25*(2), 231-243.

Youtz, Adella Clark. (1941). An experimental evaluation of Jost's laws. *Psychological Monographs, 53*(1), i.

Zimmerman, Joel. (1975). Free recall after self-paced study: A test of the attention explanation of the spacing effect. *The American Journal of Psychology*, 277-291.