Persistence and plasticity in the human memory system: An empirical investigation of the overwriting hypothesis

Tom E. Hardwicke

Department of Experimental Psychology

University College London

Thesis submitted for the degree of

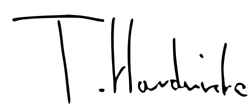
Doctor of Philosophy

August, 2016

DECLARATION

I, Tom E. Hardwicke, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

To the greatest possible extent I have endeavoured to uphold the principles of openness and transparency in this work. In order to enhance reproducibility, the thesis was written with R Markdown and LaTeX, such that the analysis code is interleaved with the main text ('under the hood'). Where possible, experimental materials, data, and analysis code pertaining to the thesis have been made publicly available on the Open Science Framework (https://osf.io/rxtgs).

SIGNED: 

Tom E. Hardwicke

London, August, 2016

To my parents, who encouraged me to Never Stop Asking Questions, and then gave me the means to try and answer some of them.

ACKNOWLEDGEMENTS

> "Finally, from so little sleeping and so much reading, his brain dried up and he went completely out of his mind."
>
> — Miguel de Cervantes, *Don Quixote*

It has been a long haul from when that lonely text cursor blinked at me repeatedly on the vast empty expanse of Thesis, page 1. But I've made it this far at least, and I am pleased to report that my brain has not dried up, and as far as I am aware, I have retained most of my principal faculties. For this I am indebted to a number of wonderful people, many of whom have waited patiently for me to emerge from several months of self-imposed isolation in my parents' garden shed.

Firstly, I would like to express my gratitude to my supervisor David Shanks. David, I feel enormously privileged to have had this opportunity to work with you these past few years. I have cherished the intellectual freedom throughout and I am so grateful to you for your patience, insight, and mentorship. Thank you!

To my colleagues and friends from the lab at UCL, Miguel Vadillo, Rosalind Potts, Maarten Speekenbrink, Manos Konstantinidis, Hannah Tickle, Simone Malejka, Eric Schulz, Chunliang Yang, and Tina Huang, thank you for sharing your brilliant minds with me in various lab meetings, coffee shops, and corridors, and of course - making 204c a fun place to be!

Miguel, I'm grateful to you for your support and encouragement, particularly when things get meta. I've learned a great deal from you about being a good scientist. Manos, you and I founded many noble flamingo and potato-based traditions over the years and I think future generations can only look back on all that as something of a Golden Era. Hannah, I'm sure we must have enjoyed an unprecedented number of cups of tea together (kudos Steve The Elephant), and yet still, I can't wait for the next one.

To my great friends, fellow PhD companions, and parliament of owls, Kari Vyas, Leila Jameel, and Matthew Jones, I'm so proud of all we achieved together, especially CRITUCL and our various open science endeavours. See you in the Alley Cat?

Thank you to Brian Nosek for allowing me to invite myself over to work with him and his team at the Center for Open Science last year. David Mellor, Billy Hunt (cheers!), Erin Braswell, and Katy Cain, thank you for making me feel so welcome such a long way from home. Who knew 'Foam Henge' was a thing?

I am grateful to the hardworking students who have assisted with data collection for some of the experiments reported in this thesis: Mahdi Taqi (Chapter 3), Gudmundur Haraldsson (Chapter 4) and Bethany Manning (Chapter 6). Best of luck to you all on your next adventures!

Several organisations have generously supported my research. The Economic and Social Research Council provided my PhD studentship, and funded my trip to COS in the United States. The Experimental Psychology Society, The Department of Experimental Psychology (UCL), and The School of Life and Medical Sciences (UCL) have all provided funding to support my attendance at various academic conferences.

Finally, to my mum, dad, and sister, I simply could not have done this without your love and support. You have been so generous and selfless, and the opportunities I have had are down to you. Thank you!

PUBLISHED AND SUBMITTED ARTICLES

Experiments 1 - 7 appear in Hardwicke, T. E., Taqi, M., & Shanks, D. R. (2016). Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proceedings of the National Academy of Sciences of The United States of America*, *113*, 5206-5211.

Experiments 1 - 7 are also discussed in Hardwicke, T. E., & Shanks, D. R. (2016). Reply to Walker and Stickgold: Proposed boundary conditions on memory reconsolidation will require empirical verification, *Proceedings of the National Academy of Sciences of The United States of America*. *113*, 1-2.

Experiments 8 - 10 appear in Hardwicke, T. E., Haraldsson, G. D., & Shanks, D. R. (under revision). Revisiting the misinformation effect: Does disruption of reconsolidation enable memory trace overwriting?

Abstract

The human memory system must resolve a critical tension, ensuring that knowledge endures over time (*persistence*), whilst simultaneously retaining a capacity for updating when knowledge is outdated or erroneous (*plasticity*). In this thesis, I examine the provocative idea that memory traces can be overwritten with new information, especially during transient periods of retrieval-induced plasticity that occur when a trace undergoes *reconsolidation*. A systematic review of human reconsolidation studies finds that the evidentiary support for this claim is remarkably tenuous. Furthermore, the theory fails to survive several strong empirical tests. In Experiments 1-7, I do not replicate a previous finding that is widely cited as a convincing demonstration of human reconsolidation. In Experiments 8-10, I revisit the 'destructive updating' account of the classic 'misinformation effect' in the context of reconsolidation theory. These experiments show that the effect can be eliminated when an appropriate recognition test is used, demonstrating that event traces are not irrecoverably lost, and therefore cannot have been overwritten during reconsolidation. In Experiment 11, I examine whether prior retrieval will help or hinder the correction of naturally occurring semantic misconceptions. Contrary to reconsolidation theory, I find that knowledge updating is not contingent on memory retrieval, nor does it result in the overwriting of prior knowledge. Finally, in the context of media 'breaking news' reports (Experiment 12), I find that the provision of an explicit retraction message, coupled with an alternative account with high causal coverage, is insufficient to eliminate reliance on false information. Finally, I contend that the widespread proliferation of *ad hoc* hypotheses, and the absence of systematic direct replication, has caused the field of reconsolidation to descend into a theoretical quagmire. I make several recommendations based on the principles of open science that may help to restore mechanisms of self-correction and foster genuine theoretical progress.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> "There is no such thing as forgetting possible to the mind; a thousand accidents may, and will interpose a veil between our present consciousness and the secret inscriptions on the mind; accidents of the same sort will also rend away this veil; but alike, whether veiled or unveiled, the inscription remains for ever."
>
> — Thomas De Quincey, 1821

We begin our lives as naïve biological entities precariously situated in a world that we do not understand. The medley of information that surrounds us is complex and overwhelming, and yet there is also a semblance of structure and order that we can exploit to our advantage. We begin to extract these regularities and imprint them, however imperfectly, upon our minds, and eventually, over the course of our lifetimes, we accumulate vast knowledge of people, places, events, and objects that we can identify, locate, anticipate, and interact with.

So deceptively effortless are these abilities that it rarely occurs to us to consider the challenges that must be overcome in order to achieve them. Learning is rarely a straightforward process because our surroundings do not remain static whilst we methodically parse and encode them. The world is constantly in flux: forests grow, rivers divert, climates change, words become archaic, politicians resign, and technologies fall into obsolescence. As a result, information stored in memory can quickly become erroneous or outdated, disrupting our ability to calibrate our behaviours to the demands of the environment. An adaptive organism must therefore retain a capacity for updating its knowledge in order to ensure that its internal representations are an accurate reflection of the external world (R. A. Bjork, 1978; Dudai, 2009; Kraemer & Golding, 1997).

Although the need for memory updating is clear, the mechanism that underlies it is not. It might seem ideal that whenever existing memory representations conflict with new information in the environment we should simply replace the former with the latter, a process analogous to a computer overwriting information on its hard drive. However, a recent change in the environment may not necessarily be representative of more global regularities that emerge across longer time scales (Speekenbrink & Shanks, 2010). Local inconsistencies (‘noise’) could easily obscure useful patterns (‘signals’) that can only be detected when we gradually accumulate information over time. Consider for example, the features of our natural environment that reliably vary with the seasons, such as fluctuations in temperature, the presence of berries in a hedgerow, or migrating swallows in the sky. It would certainly be tiresome if we had to begin each summer anew by relearning that shorts and t-shirts are more appropriate outerwear than scarfs and woolly hats. In the dynamic urban societies many of us inhabit, we frequently need to retain information that loses its utility on a short-term basis. For instance, road works may necessitate a temporary detour from our regular commute. Often knowledge that initially appeared inaccurate may later turn out to be useful once again. For the scientist, it would be inconvenient to say the least if we completely revised our beliefs based on evidence presented in a new paper, only to find out later that the study was deeply flawed. In light of such events, it would appear sub-optimal or even disastrous to simply overwrite existing knowledge whenever our internal memory representations hold some discrepancy with the external world. Consequently, the memory system must resolve a critical tension, ensuring that memory traces endure in order to preserve information accumulated over a lifetime (*persistence*), whilst simultaneously maintaining a capacity for updating when existing knowledge is genuinely erroneous or outdated (*plasticity*)[1].

The question of how we update is inherently interwoven with the question of how we forget. Although we often view forgetting as a troublesome phenomenon, many theorists recognise

[1]The terms *plasticity* and *persistence* are used across multiple levels of analysis (Bruer, 2007). Broadly speaking, persistence refers to the continued retention of information in the memory system, whereas plasticity refers to the capacity for an organism to change. Both persistence and plasticity can have specific definitions in particular contexts. For example, an experimental psychologist might investigate *behavioural plasticity* and a neuroscientist might investigate *synaptic plasticity*. Both situations broadly refer to the capacity for an entity to change, however the two situations are not necessarily synonymous: it is not (logically) inevitable that behavioural plasticity is determined by synaptic plasticity. Unless otherwise noted, the terms will be used in the broad descriptive sense throughout this thesis.

that it could have adaptive value as a mechanism for memory updating, diminishing the influence of old knowledge in order to make way for the new (R. A. Bjork, 1978; Dudai, 2009; Kraemer & Golding, 1997). However, for such a ubiquitous phenomenon, there is little consensus as to what forgetting actually entails. Colloquially, we use the term 'forget' to refer to an inability to recall something that we have previously learned. But this definition makes no commitment as to the fate of the old information. Some have suggested that memory updating is a destructive process that involves the overwriting of existing memory traces with new information (Dudai, 2009; Hardt, Einarsson, & Nader, 2010; J. L. Lee, 2009; Loftus, 1979b). For example, Nader, Hardt, Einarsson, and Finnie (2013) suggest that 'memory may be maintained in a permanently modifiable form that permits the updating of each trace when relevant new information is encountered.' This is forgetting in its strongest form: the permanent loss of information from the memory system. However, others argue that forgetting is, in a sense, illusory. When information does not come to mind, this does not necessarily require that it has been permanently lost, and could instead reflect a problem (which may be temporary) retrieving that information from the memory store (R. A. Bjork & Bjork, 1992; Hintzman, 1986; Mensink & Raaijmakers, 1988; Morton, Hammersley, & Bekerian, 1985). Furthermore, because responses on a memory test can be influenced by a number of factors unrelated to either storage or retrievability, behavioural performance cannot be taken as a direct indicator of the contents of memory (Bouton & Moody, 2004; Cahill, McGaugh, & Weinberger, 2001; M. J. Watkins, 1990). Empirical investigation of this topic has provoked a historic debate spanning multiple fields of enquiry, experimental paradigms, and species. The crux of the issue is this: when we fail to remember, has the memory trace really been lost forever?

In this thesis, we will examine the claim that memory updating is achieved through a process of overwriting. Specifically, we will focus on the theory of memory *reconsolidation* that has initiated considerable excitement in recent years. According to reconsolidation theory, retrieval of existing memory traces induces a state of instability from which the trace must re-stabilise (or 'reconsolidate') in order to persist. The time taken for this reconsolidation process to complete represents a transient period of plasticity during which the trace is open to modification. As such, the accuracy and relevance of memory traces could be routinely maintained through exposure to new environmental input that overwrites existing information during repeated 'reconsolidation windows'. Reconsolidation, therefore, may be a process that is capable of ensuring both the persistence and plasticity of the memory trace. In subsequent chapters, we will interrogate this theory through a systematic evaluation of the extant evidence used to support it, and through empirical investigations designed to provide strong tests of its predictions. We begin in the present chapter with a review of the enduring endeavour to elucidate the mechanisms of forgetting, updating, and overwriting.

1.1 A descriptive framework for investigations of memory updating and forgetting

Throughout this thesis we will use a simple descriptive framework (or "stage analysis"; Crowder, 1976) to guide our investigation of persistence and plasticity in the human memory system (Figure 1.1). The framework is loosely based on a model called the General Abstract Processing System developed by Endel Tulving (Tulving, 1983b) and related ideas originally proposed by Richard Semon (Semon, 1921; see Schacter, Eich, & Tulving, 1978). The purpose of its inclusion here is solely to create a shared frame of reference for writer and reader, not to provide formal theoretical predictions.

A common goal of the memory researcher is to infer from their subject's overt behaviour what information has been learned and retained from the environment. The framework illustrates that this is not a straightforward task because a number of intermediary factors prevent any direct access to the contents of memory. Starting at the top of the diagram (and ignoring the coloured shapes for the moment), we initially assume that some information in the environment has been perceived and encoded by the organism such that it now exists in a form of electrophysiological activity within the circuitry of the brain. We will call this state 'ecphoric memory' for reasons that will soon become apparent. For the information (in whatever representational form it now exists) to persist in memory, some *storage* processes are required to make a physical impression in the nervous system known as a *memory trace* (Bower, 1967; Tulving & Watkins, 1975; Semon, 1921; Underwood, 1969). Although the precise nature of the trace (or 'engram') is largely unknown, assuming its existence is necessary to account for how information we have learned in the past can influence us in the

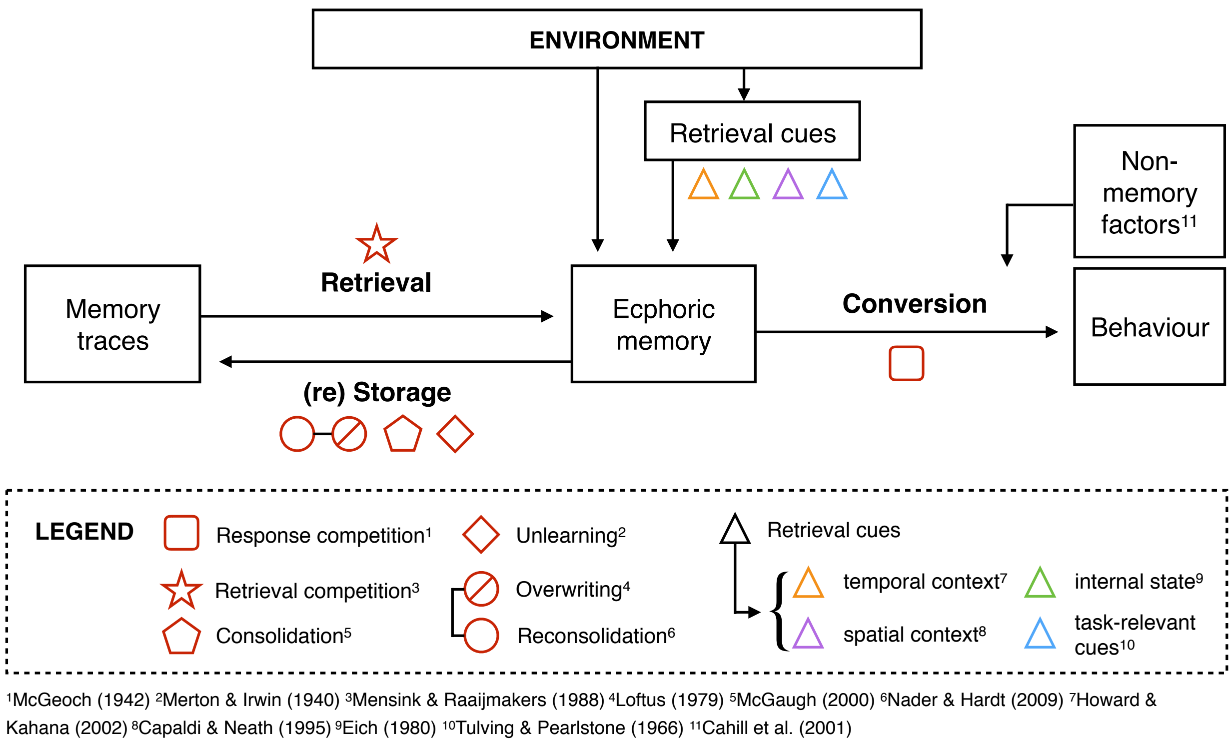


Figure 1.1: A descriptive framework for investigations of memory updating. Coloured symbols represent the loci of factors that could potentially cause forgetting. See main text for details.

present. Candidate storage processes will be discussed in greater detail later in this chapter, but they are generally thought to involve a cascade of molecular events that eventually lead to structural changes at the level of neural synapses (Dudai, 2002; McGaugh, 2000).

For the contents of memory to influence behaviour, latent memory traces must be activated through a process of retrieval (Lewis, 1979; Tulving, 1983b). Retrieval is assumed to be dependent on the presence of sufficient internal and/or external retrieval cues (Spear, 1973; Tulving, 1974, 1983b). A retrieval cue is thought to be effective because it shares some similarity (or "resonance", Semon, 1921; see Schacter et al., 1978) with the informational content of any given memory trace (Semon, 1921; Tulving, 1983a). Critically therefore, a given cue could be capable of activating multiple traces. Some retrieval cues have direct relevance to the task an individual is performing. For example, an individual's free recall of miscellaneous items from a list can be improved when they are provided with the category names for those items (Tulving & Pearlstone, 1966; Tulving & Psotka, 1971). Other potential retrieval cues include the spatial (Capaldi & Neath, 1995) or temporal context (M. W. Howard & Kahana, 2002), and the internal state of the organism (e.g., mood, arousal, drug-induced state; Eich, 1980). Retrieval cues will be discussed in more detail later in this chapter.

The synergistic interaction of incoming environmental information (including retrieval cues) and memory traces, results in a state of activity known as an *ecphoric memory* (Tulving, 1974, 1983a; Semon, 1921). Although often overlooked, this distinction between a latent memory trace and an active (ecphoric) memory has appeared in the writing of several theorists (Hintzman, 1986; Lewis, 1979; Moscovitch, 2007; Tulving, 1983a; Semon, 1921), and cautions us against the assumption that recollective experience can provide a direct and undiluted indication of the contents of memory storage. Moscovitch (2007, p.18) captures the distinction cogently: "The engram or memory trace is the representation of an encoded event or experience. It is not yet a memory, but provides the necessary (physical) condition for memories to emerge, just as an external stimulus provides the occasion for a percept to emerge... A memory emerges when the engram interacts with retrieval cues or information derived from particular environmental conditions."

Tulving (1983a) argued that the act of retrieval alone is not sufficient for the contents of the memory trace to influence an individual's overt behaviour. He suggested that an additional stage of 'conversion' was required. Conversion does not refer to a single process, but highlights that the transition from successful retrieval to overt behaviour is not direct or inevitable. For example, Tulving (1983a) suggests that part of conversion may involve the individual making a judgement or decision based on their recollective experience. If their decision was to, say, withhold a response on a memory test, the attending researcher would be none the wiser as to whether they had learned anything or not. Additionally, a long history of empirical investigation has furnished memory scientists with a substantial list of non-memory factors, such as motivation, attention, or injury, that can readily influence participants' behaviour on a memory test (Bouton & Moody, 2004; Cahill et al., 2001). Consider for example, a young pianist in a packed concert hall who successfully retrieves the memory traces that represent her well-honed ability to play the *Moonlight Sonata*. Unfortunately, to the dismay of her audience, she fails to perform the piece with any degree of expertise. Nevertheless, it is also apparent that she is feeling extremely nervous, distracted by her mother in the front row, and has a broken finger. Clearly it would be premature to conclude that the pianist has forgotten the *Moonlight Sonata*.

In summary, the framework clarifies several basic assumptions and concepts, and highlights the plight of the memory researcher. As Thomas De Quincey alludes to in the opening quotation, a number of intermediary stages intervene between the memory trace and overt (i.e., measurable) behaviour. As such, the existence of a memory trace is necessary, but not sufficient, for the successful generation of trace-dependent performance. In the event that the expected trace-dependent performance does not occur, one cannot be entirely certain whether the target information was (1) successfully stored and successfully retrieved, but not converted into overt behaviour; (2) successfully stored but not retrieved (3) successfully stored initially, but lost in the meantime (i.e., forgotten, in the strong sense); or (4) never stored in the first place. It falls to the researcher to control, to the greatest possible extent, the multiple factors that can influence performance, in an effort to isolate the critical variables of interest. In this chapter we will use the framework as a 'pin-board' (see coloured shapes, Figure 1.1) to keep track of the key factors that have been proposed to account for the absence of trace-dependent performance on a memory test: the loci of forgetting.

1.2 The loci of forgetting

At the turn of the 19th century, a series of experiments conducted by Müller and Pilzecker (1900; see Dewar, Cowan, & Della Sala, 2007; Lechner, Squire, & Byrne, 1999) laid the foundations for a debate about the loci of interference and forgetting effects that still continues to this day. In one experiment ($\text{Experiment}_{\text{MP}}$ 32[2]; see Figure 1.2) participants learned a list (List 1) of cue-target pairs consisting of nonsense syllables (e.g., *rit-zir*, *mur-geim*… etc), then either learned a new list (List 2; Interference Condition) or did not (No Interference Condition). On a subsequent test, participants were presented with the List 1 cues and asked

[2]Throughout this thesis, specific experiments from external studies will be identified using a subscript (indicating the surnames of the first and last author, and, if necessary, the year that the study was published) in order to differentiate them from experiments belonging to this thesis.

to provide the relevant targets. The findings clearly indicated lower recall in the Interference Condition (27%) relative to the No Interference Condition (55%), an effect that Müller and Pilzecker described as 'retroactive inhibition' and what is now generally known as 'retroactive interference'.

In a second important experiment (Experiment$_{\text{MP}}$ 34; see Figure 1.2), participants learned List 2 either after a brief delay (17s; Short-Delay Interference Condition) or a long delay (6m; Long-Delay Interference Condition). Recall of List 1 was substantially more impaired after a brief delay (28%) compared to after a long delay (49%)[3]. Müller and Pilzecker (1900) attributed these effects to the disruption of continuing physiological activity necessary for the List 1 representation to become fixated in the memory system, a process they referred to as 'consolidirung' ('consolidation', from the Latin 'consolidare' meaning *to make firm*). The reduced impairment in the Long-Delay Interference Condition suggested that this consolidation process was time-dependent, and implied a brief 'consolidation window' during which a newly acquired memory trace would be vulnerable to interference. In terms of our stage analysis (Figure 1.1, red pentagon) consolidation reflects a storage process that must be allowed to resolve in order for a newly acquired memory trace to persist in the memory store.

In subsequent years, the debate about interference and forgetting appears to have fragmented across research communities (Hardt et al., 2010; Lechner et al., 1999; Wixted, 2004). Following a number of empirical investigations by John McGeoch (J. A. McGeoch & McDonald, 1931; J. A. McGeoch, 1933a, 1942), many psychologists pivoted away from the consolidation hypothesis to focus on other potential causes of retroactive interference, including *unlearning* (Melton & Irwin, 1940) and *response competition* (J. A. McGeoch, 1932, 1942). By contrast, consolidation theory remained highly influential within neuroscientific circles (and amongst biologically-oriented psychologists), stubbornly resisting several challenges during the 1960s (cf. Lewis & Maher, 1965, 1966; McGaugh & Petrinovich, 1966) and early 1970s (cf. McGaugh & Dawson, 1971; R. R. Miller & Springer, 1973), and eventually emerging as the dominant

[3]Ideally this experiment should also have included 'No Interference Conditions' to ensure the effect could be attributed to the interaction of delay and interference rather than a just an effect of delay. However, performance in the Long-Delay Interference Condition appears to be comparable to the No Interference Condition of Experiment$_{\text{MP}}$ 32.

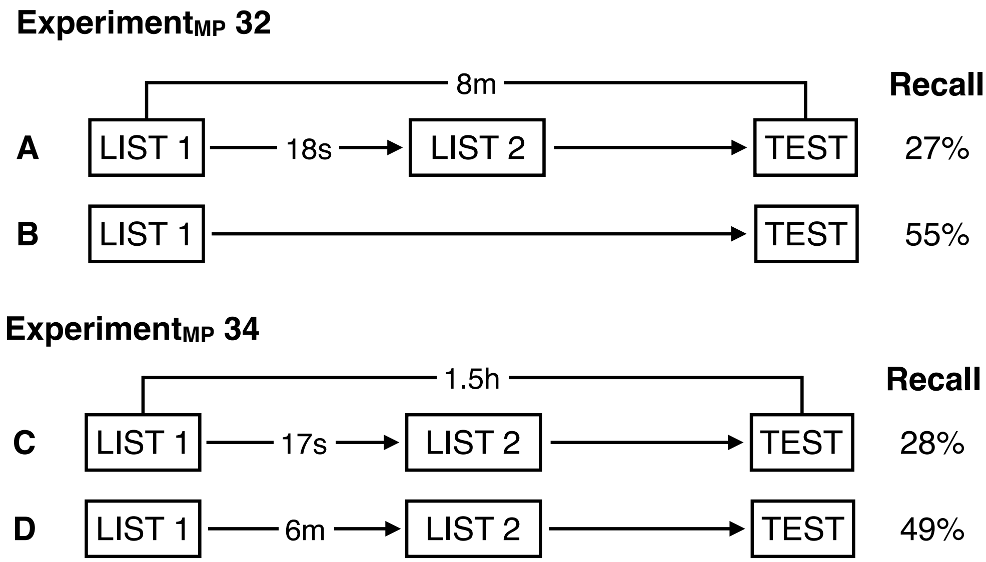


Figure 1.2: Design of Experiment$_{\mathrm{MP}}$ 34 in Müller and Pilzecker (1900; see Lechner et al., 1999). **A**: Interference Condition; **B**: No Interference Condition; **C**: Short-Delay Interference Condition; **D**: Long-Delay Interference Condition. See main text for details.

neurobiological explanation of amnesia (Dudai, 1996; McGaugh, 2000; Wixted, 2004).

We will now follow each divergent path of the two research communities[4] in turn, first tracing the arc navigated by many psychologists, and then returning to the arc predominantly taken by more biologically-oriented researchers. Towards the end of the chapter, we will see these two arcs converge: at the turn of the 20th century, 100 years after Müller and Pilzecker first sparked interest in consolidation, the new theory of memory *reconsolidation* dramatically took center stage in the neuroscientific community (Dudai, 2000; López, 2000; Nadel & Land, 2000; Nader, Schafe, & LeDoux, 2000b; Sara, 2000b) and offered to bridge the historical divide with psychology (Hardt et al., 2010). That is where the empirical investigation reported in this thesis begins.

[4]Naturally there was some degree of cross-talk between disciplines (e.g., Crowder, 1976; Melton, 1963; Spear, 1973) but the theoretical and empirical zeitgeists can clearly be differentiated (Lechner et al., 1999; Hardt et al., 2010; Wixted, 2004).

1.2.1 Response competition

Many studies that followed Müller and Pilzecker (1900) adopted a similar procedure that became known as the paired-associates paradigm (Figure 1.3, for review see M. C. Anderson & Neely, 1996; Crowder, 1976). During the Training Phase, participants learned a list of multiple cue-target word pairs (the standard notation to refer to these items is A-B, where A refers to the cue and B refers to the target). Horse – Leaf, Car – Pizza, or Baseball – Fountain could all be examples of A-B pairs. During a subsequent Test Phase, participants were provided with the cues (A), and ask to recall the targets (B). For example, when presented with the cue 'Horse', the correct response would be 'Leaf'. This type of test was known as 'Modified Free Recall' (MFR, Underwood, 1948) because unlike standard free-recall, the participant was directed to give a specific response (target B) to a specific cue (A).

A critical feature of this paradigm was the nature of an interpolated New Learning Phase that fell between the Training and Test Phases. During this New Learning Phase, participants were asked to memorise a second list of cue-target pairs which often shared components of the first list. For example, an A-C list would include the cues (A) from the Training list paired with new targets (C), for example, Horse – Chair. Alternatively, a researcher might use different cues and different targets (C-D), for example, Mug – Brick. Typical control groups required participants to complete a distractor task, asked them to 'rest', or simply involved participants skipping the New Learning Phase entirely.

The consolidation account began to look unsatisfactory after a series of studies did not obtain the predicted temporally graded decline in retroactive interference (Archer & Underwood, 1951; J. A. McGeoch, 1933a, 1933b; J. A. McGeoch & Nolen, 1933; Newton & Wickens, 1956; E. S. Robinson, 1920). In addition, interference effects could be obtained when the interpolated material was introduced after an especially long delay, for example 6 weeks. It did not seem likely that organised patterns of neural activity activated during the original training episode would persevere over such a long period of time (J. A. McGeoch, 1942). Finally, the data indicated that the magnitude of interference varied as a function of the similarity of the original and interpolated learning material, a phenomenon not anticipated by consolidation theory (J. A. McGeoch & McDonald, 1931; J. A. McGeoch & McGeoch, 1937). We will

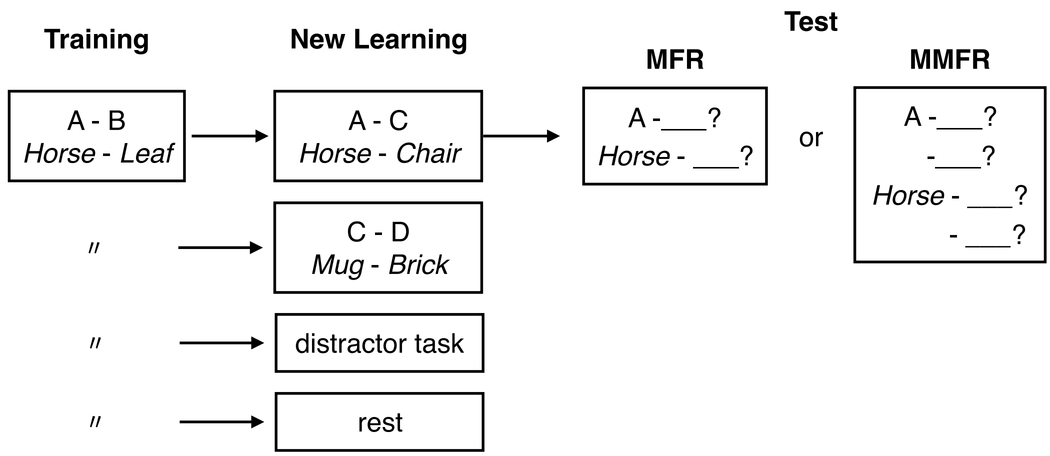


Figure 1.3: Designs of typical paired-associates paradigms. Letters (A, B, C, D) represent the standard notation used to refer to cue-target pairs. Words in italics represent examples of cue-target pairs. *MFR* = Modified Free-Recall test. *MMFR* = Modified Modified Free-Recall test. See main text for details.

examine both conclusions in more detail later in this chapter (see subsubsection 1.2.7.4), but at the time these findings appeared to provide a compelling case against consolidation theory and contributed to the field turning its focus to new explanations of the retroactive interference phenomenon.

To account for the findings, McGeoch proposed that retroactive interference was "a function of competition among responses, with a resultant momentary dominance... of one response over another. Responses... are not necessarily lost from the subject's repertoire, but are kept by other responses from appearing" (J. A. McGeoch, 1942, p. 495). This response competition theory explicitly excluded a storage-based forgetting process and suggested that A-B and A-C associations coexisted independently in the memory store (also see Martin, 1971). In terms of our stage analysis (Figure 1.1) the contrast between consolidation theory (red pentagon) and response competition theory (red square) is stark. Consolidation theory suggests that post-acquisition new learning disrupts the stabilisation of memory traces, preventing the adequate storage of target information. Conversely, response competition theory suggests that the target information persists in memory storage, and is successfully retrieved, but may not be expressed in overt behaviour because the newly acquired response is dominant. Furthermore, whereas consolidation theory implies that interventions will only be effective during a brief post-acquisition 'consolidation window', response competition theory suggests that new learning can potentially cause interference at any time, assuming it bears some similarity to target information stored in memory.

The A-B, A-C paradigm provides an empirical model of a simple real-world memory updating scenario. For example, consider an individual who learns that in situation A (e.g., using a cash machine), they will be best served by performing response B (e.g., entering the pin number '1234'). However, the external world changes: following a series of security breaches the bank decides to issue all of its customers with new pin numbers (C). The individual must now learn to associate situation A with response C (e.g., entering the pin number '5678'). Would this also lead to forgetting of response B? McGeoch's theory suggests that both A-B and A-C associations will co-exist in the memory store but one response (say C) will come to out-compete the other response (say B) at the point of retrieval. In such a scenario, the apparent forgetting of response B would be illusory.

1.2.2 Unlearning

McGeoch's theory made (at least) two important empirical predictions: (1) because retroactive interference is caused by List 2 items blocking List 1 items on the recall test, an increase in retroactive interference should be accompanied by a concomitant increase in the number of intrusion errors (List 2 items being provided instead of List 1 items); (2) because retroactive interference is the result of the momentary dominance of List 2 items over List 1 items, it should be alleviated if participants are allowed to provide both responses. The first prediction was found wanting in a study by Melton and Irwin (1940) and the second in a study by Barnes and Underwood (1959).

In Melton and Irwin (1940; also see a partial replication using an A-B, A-C paradigm by Thune & Underwood, 1943) participants ($n = 24$) learned a series of lists consisting of 18 nonsense syllables each presented for 5 repetitions. They then learned a series of second lists prior to test. The number of repetitions of the second list was manipulated within-subjects such that there were either 0 (i.e., 'rest'), 5, 10, 20, or 40 repetitions. The findings are shown in Figure 1.4. Retroactive interference became more pronounced with an increasing

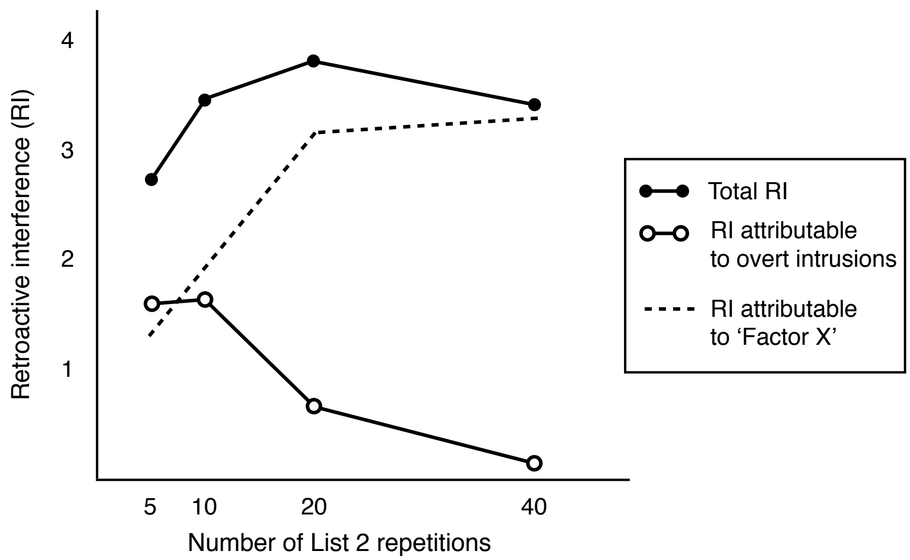


Figure 1.4: Findings of Melton and Irwin (1940). Mean retroactive interference (RI, i.e., the difference between performance in a resting control group relative to the experimental groups) as a function of degree of interpolated learning. Adapted from Crowder (1976).

number of List 2 repetitions (black circles). However, after a moderate degree of interpolated learning, the increase in interference was not accompanied by a corresponding increase in the frequency of List 2 intrusions (white circles). After accounting for the retroactive interference that could be caused by List 2 intrusions, Melton and Irwin were left with an unexplained performance impairment (dashed line). The researchers proposed a two-factor theory comprising of McGeoch's response competition (to account for interference that could be attributed to intrusions) and a mysterious 'Factor X' (to account for interference that could not be attributed to intrusions). Tentatively, they suggested that Factor X could be an 'unlearning' mechanism that involved weakening of the List 1 information: "It may be that there is a direct weakening of the original responses at the time the interpolated responses are being learned, and that this unlearning factor accounts for the discrepancies noted" (Melton & Irwin, 1940, p. 197). Therefore, two-factor interference theory consisted of both conversion-based (response competition, Figure 1.1, red square) and storage-based (unlearning, Figure 1.1, diamond) forgetting mechanisms (for review see Postman & Underwood, 1973).

In a subsequent study, Barnes and Underwood (1959) also sought evidence for the 'fate' of the original learning material in an A-B, A-C paradigm. The researchers noted that in

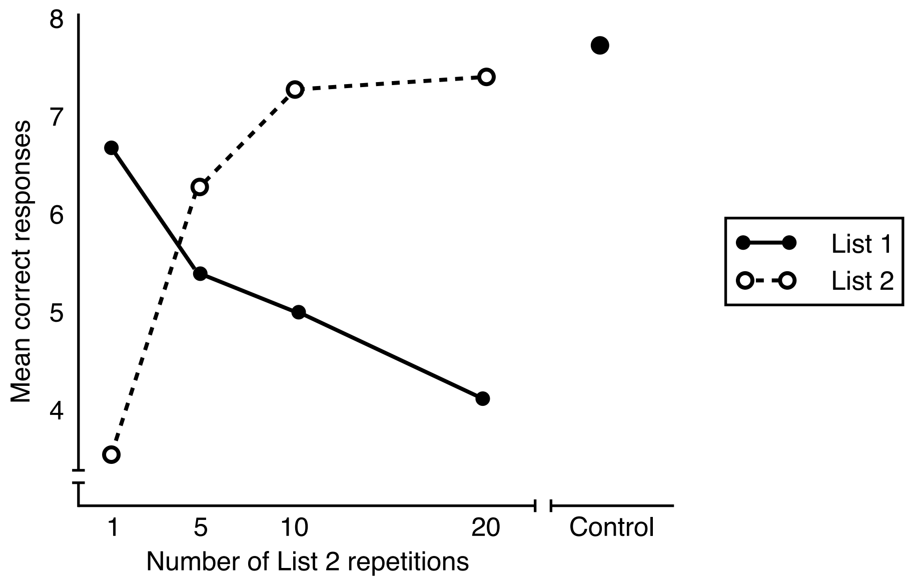


Figure 1.5: Findings of Barnes and Underwood (1959). Mean correct responses on an MMFR test in an A-B, A-C paired associates learning paradigm. In the control condition, participants did not learn List 2 and instead completed puzzles during the delay between training and test. Adapted from Barnes and Underwood (1959).

previous experiments (e.g., Underwood, 1948; Briggs, 1954) participants were only asked to provide a single response for each item at the test stage (an 'MFR' Test, see Figure 1.3). Therefore, it could be the case that both original and interpolated items are successfully retrieved by participants, but their behaviour only reveals the most dominant amongst competing responses. Barnes and Underwood (1959) altered the standard MFR test by asking participants to provide *both* responses (i.e., B and C in an A-B, A-C paradigm) rather than just a single response. This adjustment was critical because it could eliminate the contribution of response competition. If both responses are available to the participant (i.e., successfully retrieved), it is irrelevant whether one is dominant over the other when they can both be provided on the test. This type of test entered into regular use, and earned the imaginative name 'modified modified free-recall' (MMFR, see Figure 1.3, Postman & Underwood, 1973). Barnes and Underwood (1959) found (Figure 1.5) that as the degree of interpolated learning increased, List 2 recall also increased but List 1 recall declined. Because the role of response competition had been removed, this was taken to indicate compelling support for the role of unlearning.

Nevertheless, evidence that seems inconsistent with unlearning began to emerge as several studies observed retroactive interference effects diminishing with the passage of time. Such 'spontaneous recovery' effects have been reported on numerous occasions in the paired-associates paradigm using the MFR (e.g., Briggs, 1954; Underwood, 1948) and MMFR tests (e.g., Forrester, 1970; Silverstein, 1967), although in some studies the effects were small or absent (e.g., Birnbaum, 1965; Koppenaal, 1963; Slamecka, 1966). An early review of the literature indicated that the presence or absence of a recovery effect was to some extent dependent on the researchers' method of statistical analysis (A. S. Brown, 1976), and a more recent study Wheeler (1995) employing a within-subjects design observed robust spontaneous recovery effects across a range of task parameters. Because spontaneous recovery effects imply that the original material is at least partially intact in the memory store, the unlearning hypothesis cannot provide a full account of forgetting in the A-B, A-C paradigm.

1.2.3 Extinction, counter-conditioning, and recovery in non-human animals

Interestingly, some researchers, such as Underwood (1948), regarded cases of spontaneous recovery as evidence *for* unlearning rather than against it. They noticed that the phenomenon of 'extinction' widely reported in the non-human animal literature, provided a useful analogy for retroactive interference effects observed in the A-B, A-C paradigm (Underwood, 1948). However, this explanation by analogy does not provide evidence for unlearning because extinction itself could have multiple theoretical interpretations. The concept of unlearning did have some influence in the non-human animal literature as a candidate account of extinction (for reviews of translational research in humans see Bouton, 2014; Haaker, Golkar, Hermans, & Lonsdorf, 2014; Vervliet, Craske, & Hermans, 2013). Most prominently, an unlearning process was formally instantiated in the influential Rescorla-Wagner model as a decrease in associative strength between two stimuli (Rescorla & Wagner, 1972; for review see R. R. Miller, Barnet, & Grahame, 1995). However, as outlined below, extensive empirical investigation strongly suggests that extinction does not involve storage-based memory impairment (Bouton, 2002).

Extinction is most clearly demonstrated in the Pavlovian conditioning paradigm (Bouton

& Moody, 2004; Pavlov, 1927). The researcher first identifies a motivationally significant event known as the 'unconditioned stimulus' (US, e.g., a foot shock) which reliably elicits a reaction from the animal known as the 'unconditioned response' (UR, e.g., a fear response). During the Training Phase, a neutral event known as the 'conditioned stimulus' (CS, e.g., a tone) is presented in conjunction with the US, often on repeated trials (although learning is sometimes observed after a single trial). If conditioning has been successful, the animal will exhibit the UR (now called the 'conditioned response', CR) when the CS is presented on its own, implying that the animal has learned an association between the CS and the US (e.g., tone and foot-shock). If the animal is now repeatedly presented with the CS on its own (Extinction Phase), performance typically declines to the point where the CS no longer reliably elicits the CR.

It is tempting to conclude that the conditioning process has been reversed and the association between the CS and the US 'unlearned' (Rescorla & Wagner, 1972). Indeed, the name extinction implies destructive loss. However, a considerable body of evidence suggests that the original CS-US association is (at least partially) intact (for reviews see Bouton, 2002; Falls, 1998; Rescorla, 2001). Firstly, extinction effects appear to be context-dependent. If the animal receives Training trials in Context A, and then Extinction trials in Context B, responding to the CS can be restored if the animal is returned to Context A for a Testing Phase ("renewal", e.g., Bouton & Bolles, 1979a; Bouton & King, 1983). Secondly, it has been observed that reminder presentations of the US following the Extinction Phase can restore responding to the CS ("reinstatement", e.g., Bouton & Bolles, 1979b; Rescorla & Heth, 1975). It may also be significant (see below) that reinstatement effects tend to be dependent on the reminder US being presented in the same context as used during the Training and Testing Phases (Bouton & Bolles, 1979b; Bouton & King, 1983). Finally, it has been observed that the mere passage of time can allow for restoration of responding to the CS (Pavlov, 1927; Robbins, 1990), a 'spontaneous recovery' effect akin to that observed in the paired-associates learning paradigm (A. S. Brown, 1976; Wheeler, 1995).

Recovery effects are also repeatedly observed in counter-conditioning studies (CS-US1, CS-US2) which more closely resemble the structure of the paired-associate learning paradigm

(A-B, A-C) relative to extinction studies (CS-US, CS-noUS). In a typical counter-conditioning study, the animal might first (Training Phase) experience pairings of a tone (CS) with a foot shock (US1) and then (Transfer Phase) experience pairings of the same tone (CS) with food (US2). By the end of the Training Phase, the CS elicits an aversive response (e.g., 'freezing' in anticipation of shock). However, by the end of Transfer Phase the CS instead elicits an appetitive response (e.g., 'headjerks' in anticipation of food). Based on this outcome alone, one might be led to conclude that an association between CS-US1 formed during the Training Phase had been overwritten by a new CS-US2 association formed during the Transfer Phase. However, as with extinction studies this does not appear to be the case. The fear response can be 'renewed' if the context changes from Training (Context A), to Transfer (Context B), and reverts for Testing (Context A, Peck & Bouton, 1990). Similarly, the fear response can be 'reinstated' if rats are exposed to US1 in the same context as the subsequent Testing Phase (D. C. Brooks, Hale, Nelson, & Bouton, 1995). Finally, spontaneous recovery can be observed when there is a delay (e.g., 28-days) between the Transfer and Test Phases (Bouton & Peck, 1992).

How can these recovery effects be accounted for? One intriguing proposal is that the physical and temporal context (Figure 1.1, purple triangle and orange triangle respectively) act as special retrieval cues that 'set the occasion' for behavioural responding (Bouton, 1993; R. R. Miller & Oberling, 1998). The idea is that by the Final Test Phase the initial cue (A or CS) has become associated with two different 'meanings' (for paired-associates: B and C; for extinction: US and noUS; for counter-conditioning: US1 and US2). Consequently, the cue alone only provides an ambiguous signal about the behaviour that the animal should perform. The context 'sets the occasion' in the sense that it disambiguates the cue, and signals to the animal which of the competing responses is the most appropriate in the present situation.

In the case of the renewal effect for instance (see above), the animal appears to learn that in Context A (Training Phase), the CS signals the US (or US1 for counter-conditioning); whereas in Context B (Extinction/Transfer Phase), the CS signals noUS (or US2 for counter-conditioning). A similar explanation can be applied to reinstatement effects which also appear to be context-dependent (see above). Following the Extinction/Transfer Phase, Context A indicates that the CS signals noUS (or US2). However, the reminder presentation of the US (or US1) in Context A indicates to the animal that the CS now signals the US (or US1) again.

Spontaneous recovery can also be considered a specific instance of the renewal effect that focuses on temporal rather than physical context. According to this account, the association formed during the Extinction/Transfer Phase is also associated with a specific temporal context. As the delay until the Testing Phase increases, the similarity to the temporal context of Extinction/Transfer decreases, reducing the likelihood that the CS-noUS/CS-US2 association will be retrieved (Bouton, 1993). Clearly however, the same situation applies to the temporal context associated with the original CS-US/CS-US1 association formed during Training. Similarly, although renewal effects are most robust when the Training and Testing contexts are matched (i.e., ABA), they are also observed when the Testing context is neutral (i.e., ABC, see J. A. Harris, Jones, Bailey, & Westbrook, 2000). Therefore, the 'occasion setting' account requires us to assume that the Training association is considered as the 'default' behaviour associated with the cue, allowing it to generalise more freely across contexts. Conversely, the Extinction/Transfer association is considered an exception, and therefore more dependent on contextual cues (see Bouton, 2002).

1.2.4 Cue-dependent forgetting in humans

The evidence from non-human animal studies reviewed above clearly indicates an important role for various retrieval cues in the phenomenon of forgetting and interference. They also indicate that memory updating does not necessarily involve storage-based loss of information, and can be achieved via retrieval-based mechanisms. In many situations that might otherwise suggest the loss of information from memory, provision of appropriate retrieval cues reveals that the target memory traces have persisted. The role of retrieval cues has also received considerable attention in the human literature. In fact John McGeoch, originator of the response competition theory, also recognised that retrieval cues have an important role: "forgetting, in the sense of functional inability or loss, may result from a lack of the proper eliciting stimulus" (J. A. McGeoch, 1932, p. 365). McGeoch's evidence at this point was entirely anecdotal, and requires some historical context to appreciate: "The missionary, after being for some time in this country, loses his command of Chinese, but regains it, with almost no relearning, upon return to the stimulating environment in which he had learned and habitually used the language" (J. A. McGeoch, 1932, p. 366).

McGeoch's speculations anticipated some of the most important empirical developments in the loci of forgetting debate. In a series of studies, Endel Tulving and his colleagues obtained compelling evidence for the crucial role of retrieval cues in the process of remembering (e.g., Tulving & Pearlstone, 1966; Tulving & Psotka, 1971; Tulving & Thomson, 1973). In one prominent study by Tulving and Pearlstone (1966), participants were first instructed to memorise lists of words organised into semantic categories and presented auditorily. For example, under the category heading FOUR-FOOTED ANIMALS, one might hear the words COW, RAT, PIG, and HORSE. The number of words in a list (12, 24, 48), and the number of items per category (1, 2, 4) was varied in a fully crossed fashion between participants, giving rise to 9 conditions. For example, in the 12 word list, items were either all from a single category, two different categories, or four different categories. The additional, critical manipulation, was the type of testing. Participants were either given a *free-recall* test or a *cued-recall* test (consequently there were 18 conditions in total, $n \approx 52$ per condition). On the free-recall test, participants were given 2 minutes to write down as many training items as possible in whatever order they preferred. On the cued-recall test, participants were presented with each category name in turn, and given a short period of time to write down the training items belonging to that category.

The critical outcome was facilitated recall performance in the cued-recall condition relative to the free-recall condition (Figure 1.6). To account for this, Tulving and Pearlstone introduced an important distinction between the *availability* and *accessibility* of information in memory. As both cued-recall and free-recall groups were treated identically up until the point of testing, it is reasonable to assume that they had learned an approximately equivalent number of words. In other words, they had the same amount of information available in memory storage. However, the higher performance in the cued-recall groups relative to the free-recall groups suggests that the accessibility of information was facilitated by the provision of additional retrieval cues (i.e., category names; see Figure 1.1, blue triangle).

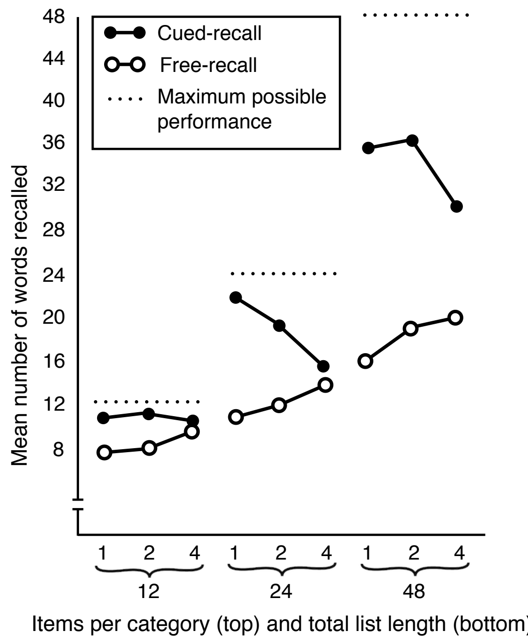


Figure 1.6: Findings of Tulving and Pearlstone's (1966) study illustrating the availability/accessibility distinction. Mean number of words recalled are shown as a function of items per category and total list length. Adapted from Tulving and Pearlstone (1966).

This experiment also illustrates another important principle related to the retrieval cue. Notice that the effectiveness of the retrieval cues appears to decline as the number of words per category increases (Figure 1.6). One way to explain these findings is in terms of the cue-overload principle (M. J. Watkins & Watkins, 1976). According to this idea, the more information (i.e., memory traces) associated with a single retrieval cue, the less effective that cue will be at facilitating access to any specific trace in memory storage[5]. Thus in Tulving and Pearlstone's (1966) experiment, when more items were associated with a category label it was less effective at facilitating retrieval of those items.

Following Tulving and Pearlstone (1966), and in parallel with work in the non-human animal literature (Spear, 1973), researchers began to explore the role of retrieval cues that were seemingly unrelated to the target information, but happened to be correlated with it. These included external cues, such as the physical context (Figure 1.1, purple triangle) and

[5]A similar concept is the 'fan effect' which refers to the idea that the more facts that are known about a particular concept, the longer the time taken to retrieve any specific fact (J. R. Anderson, 1974; J. R. Anderson & Reder, 1999).

internal cues, such as the organism's state of alertness (Figure 1.1, green triangle). In one study demonstrating the role of physical context (for review see Capaldi & Neath, 1995; S. M. Smith & Vela, 2001), Godden and Baddeley (1975) observed a remarkable interaction between the learning and testing environment. The researchers asked divers to learn a list of 36 words either on land or underwater, and then recall those words, either on land or underwater[6]. Recall was markedly improved when the learning and testing environments were congruent (dry-dry/wet-wet) compared to when they were incongruent (dry-wet/wet-dry). Studies adopting a similar design have demonstrated that the subject's internal state can also influence recall (for review see Eich, 1980). For example, in Eich, Weingartner, Stillman, and Gillin (1975) participants were asked to learn lists of words either under the influence of marijuana or placebo, and then recall those words under the influence of marijuana or placebo. The findings mirrored those of Godden and Baddeley (1975): free recall performance was facilitated when participants' internal state was congruent between learning and testing (marijuana-marijuana/placebo-placebo) compared to when it was incongruent (marijuana-placebo/placebo-marijuana).

Findings such as these led Tulving and colleagues to propose the *encoding specificity principle* (Tulving & Thomson, 1973; Tulving, 1983b), which reinforced McGeoch's ideas from four decades previously (see above, J. A. McGeoch, 1932). According to the encoding specificity principle, retrieval cues are only effective if they were encoded at the same time as the target item. More broadly, the concept of *encoding-retrieval match* (see Nairne, 2002) suggests that maximising the similarity between the encoding and retrieval conditions will facilitate trace-dependent performance. As any type of retrieval cue could contribute to this process, the concept of encoding-retrieval match could potentially explain a broad range of cue-dependent forgetting effects, including the role of context and state (Roediger & Guynn,

[6]Godden and Baddeley's rationale for using divers was apparently due to dissatisfaction with previous manipulations of environmental context which they described as "extremely artificial". Although their own study is unlikely to win any awards for ecological validity, it was certainly an improvement relative to a rather unusual study (Rand & Wapner, 1967) cited in their introduction in which the researchers "strapped their subjects to a board which was then rotated so as to keep the subject either supine or erect" (Godden & Baddeley, 1975, p. 326).

1996). However, there is an important caveat which was alluded to earlier in the context of Tulving and Pearlstone's study (see above): the cue-overload principle (M. J. Watkins & Watkins, 1976). If the effectiveness of a retrieval cue decreases as the number of traces it is associated with increases, then encoding-retrieval match cannot be the only determinant of successful retrieval.

Recognising the importance of both principles, Nairne and colleagues (Nairne, 2002; Poirier et al., 2012) have taken the position that memory retrieval is fundamentally a discrimination problem. They proposed that the apparent effectiveness of increasing encoding-retrieval similarity is largely due to the fact that it is typically associated with a concomitant increase in the cue's ability to discriminate the target trace from competitor traces. However, this will not always be the case. Consider, for example, a situation in which there are two traces, A, B, and two retrieval cues, X and Y. Cue X is uniquely associated with trace A, which also happens to be our target. Cue Y is also associated with target trace A, and associated with an additional competitor trace B. The provision of cues X and Y together would therefore maximise the encoding-retrieval match for trace A. However, because cue Y is associated with trace B, this also increases competition, or 'cue overload'. Provision of cue X alone, in contrast, yields lower encoding-retrieval match, but does not elicit competition from trace B. Therefore, according to the memory-as-discrimination hypothesis (Nairne, 2002; Poirier et al., 2012), retrieval of the target trace A will be facilitated by the provision of only cue X relative to cues X and Y together, despite the fact that in the latter case there is greater encoding-retrieval match.

Poirier et al. (2012) obtained evidence for this hypothesis in a series of experiments. In the training phase, participants learned associations between targets (e.g., nonsense syllables) and cues (e.g., symbols). Targets were associated with multiple cues, some of which uniquely predicted the target, and some of which were shared with other targets. For example, the target 'jek' might be associated with the star, square, and triangle symbols, and the target 'cef' might be associated with the oval, rhombus, and triangle symbols. In this case the triangle is a shared cue that predicts both targets, whilst the other cues uniquely predict their associated targets. In the test phase, participants were presented with various cue combinations and asked to select a single target. The critical comparison was between two conditions: when a single unique cue was presented (low encoding-retrieval match, high discriminability; $M^{L}D^{H}$) and when a unique and shared cue were presented together (high encoding-retrieval match, low discriminability; $M^{H}D^{L}$). Consistent with the memory-as-discrimination view, and contrary to the encoding-retrieval match view, response times were slower in the $M^{H}D^{L}$ condition relative to the $M^{L}D^{H}$ condition. In summary, it would seem that increasing encoding-retrieval similarity only facilitates performance if it increases the ability of the retrieval cue to discriminate between competing memory traces (Nairne, 2002; Poirier et al., 2012). Additional implications of these findings will be discussed in a subsequent section (subsection 1.2.6).

1.2.5 The misinformation effect and the overwriting hypothesis

Despite growing recognition for the role of 'retrieval-failure' as a cause of forgetting, it was not long before a new storage-based hypothesis took center stage. After decades of research dominated by list-learning experiments, Elizabeth Loftus introduced a new approach inspired by situations involving eye-witness testimony (Loftus, Miller, & Burns, 1978; Loftus, 1979b). Nevertheless, Loftus' experiments and those that followed, adhered to essentially the same core structure as the classic paired-associates paradigms of the previous era (for review see Titcomb & Reyna, 1995; Roediger & Guynn, 1996). The original 'misinformation paradigm' consisted of three phases (e.g., Loftus et al., 1978). Participants first 'witnessed' an event by viewing a series of slides depicting a traffic collision. In the second phase, some participants ('misled' group) were exposed to misinformation about the event whereas others were not ('control' group). For example, the misled group heard a reference to a 'stop sign' when the original slide sequence actually contained a 'yield sign'. By contrast, the control group heard the 'stop sign' correctly referred to as a 'stop sign' (or in some studies, a neutral reference to a 'sign'). In the final test phase, participants completed a forced-choice recognition test which included the event item and the misleading item as options. The key finding was that individuals were much more likely to respond with the misleading item when they had been exposed to misinformation compared to when they had not. Loftus concluded that the misinformation had caused the 'destructive updating' or 'overwriting' of the memory trace formed during the original event (Loftus, 1979a, 1979b; Loftus, Schooler, & Wagenaar, 1985, see Figure 1.1, red circle-slash). This striking 'misinformation effect' appeared to cast serious doubt on the veracity of eye-witness testimony (Loftus, 1979b) and sparked a fervent industry of empirical inquiry (for review see Ayers & Reder, 1998; Loftus, 2005; Zaragoza, Belli, & Payment, 2006).

Although there were clear parallels with the classic paired-associates paradigm (i.e., sign-Stop, sign-Yield ≈ A-B, A-C), these were rarely acknowledged explicitly (although see Chandler, 1989; Bowman & Zaragoza, 1989). The overwriting hypothesis seems highly similar to the unlearning hypothesis, although the latter posits a gradual weakening of memory associations whereas the former seems to imply more robust and immediate destruction. This has serious implications for theories of memory updating and it is unclear how persistence and plasticity could be effectively balanced if such a process were in operation (a point we will return to in Chapter 7).

The disconnect with the preceding interference literature could explain why relevant findings from a paired-associates study by R. C. Anderson and Watts (1971) appear to have been initially overlooked (also see Postman & Stark, 1969). Critically, R. C. Anderson and Watts showed that retroactive interference in an A-B, A-C design could be almost eliminated on a recognition test that did not include the interpolated item as a distractor (the distractors were all items from List 1). By contrast, when the recognition test did include the interpolated item as a distractor (as in Loftus et al., 1978), retroactive interference effects were obtained. This outcome provides further evidence against the unlearning hypothesis in the paired-associates paradigm, and places the loci of the interference effect firmly in the post-storage (i.e., retrieval/conversion) stages of analysis (Figure 1.1). Loftus (1979b) had suggested that the misinformation effect involved overwriting of the original event memory traces, but that conclusion would not hold if the effect were absent when a recognition test without the interpolated item is employed.

Although it was not clear if they were aware of the R. C. Anderson and Watts study, M. McCloskey and Zaragoza (1985a) conducted misinformation experiments that mirrored key aspects of their design. Firstly, the researchers replicated the misinformation effect using the same recognition test (the ‘Original Test’) employed by Loftus et al. (1978). However, when the recognition test (‘the Modified[7] Test’) excluded the misled item (i.e., the options were the event item and a novel distractor) the effect was eliminated. This study provided compelling evidence against the overwriting hypothesis because it showed that information acquired during the event was intact in the memory store. Potential reasons for the different outcomes on the two types of test and other aspects of the misinformation literature will be reviewed in more detail in a subsequent chapter (Chapter 4).

1.2.6 Retrieval competition

The research examined thus far suggests that forgetting is not necessarily attributable to storage-based factors. Apparently lost information can often be recovered through the provision of effective retrieval cues (Bouton, 2002; Tulving, 1983b), or return spontaneously with the passage of time (Rescorla, 2004; Wheeler, 1995). The finding that retroactive interference effects can be attenuated or even eliminated on appropriately designed recognition tests (R. C. Anderson & Watts, 1971; M. McCloskey & Zaragoza, 1985a; Postman & Stark, 1969) is a compelling demonstration that the original learning material has not been unlearned or overwritten.

Nevertheless, it remains unclear how to account for the retroactive interference effects observed using the MMFR test in a paired-associates learning paradigm. Note that the encoding-retrieval match hypothesis proposes that cues facilitate the *retrieval* process (Figure 1.1) by enabling access to target memory traces (Tulving, 1983b). By contrast, the occasion setting account suggests that retrieval cues facilitate the *conversion* process by helping the individual make an appropriate choice between competing responses (Bouton, 2002). The memory-as-discrimination perspective on the other hand (discussed above, subsection 1.2.4;

[7]The reader may have noticed that memory researchers have a habit of creating ‘modified’ tests. For clarity, McCloskey and Zaragoza’s ‘Modified Test’ is a *recognition* test and not be confused with the Modified Free Recall Test (MFR; Underwood, 1948), or the Modified Modified Free Recall Test (MMFR; J. M. Barnes & Underwood, 1959) which are both *cued-recall* tests.

Poirier et al., 2012), does not appear to draw a distinction between the retrieval and conversion stages. In the experiments outlined above (Poirier et al., 2012) all of the potential responses were available to the participants on a recognition test, and the manipulation of target discriminability only affected response times and not accuracy levels. As such, it is unclear whether the slower response times reflected poorer *retrieval* efficiency, or time taken actively choosing between available responses (i.e., a *conversion* issue).

These factors are not necessarily mutually exclusive, but even together they do not seem to provide an adequate explanation for interference effects on the MMFR test. Firstly, because the various parameters of the encoding and retrieval situation are typically the same for both the interference and control groups, an absence of effective retrieval cues cannot account for the apparent forgetting of B targets. Secondly, because the MMFR test allows multiple responses to be provided, and unlimited time to do so, the interference cannot be attributed to the response competition envisioned by the occasion-setting and memory-as-discrimination accounts (indeed this was the reason for introducing MMFR in the first place, see subsection 1.2.2). In earlier decades, a satisfactory explanation continued to evade theorists and the two-factor interference theory fell into “a state of ferment if not disarray” (Postman, 1975, p. 327; also see Tulving & Madigan, 1970).

Modern interference-based theories however, have fallen back on a refined version of J. A. McGeoch’s response competition principle, and dispensed with the unlearning assumption. The ‘Search of Associative Memory’ (SAM, Mensink & Raaijmakers, 1988) model for instance, emphasises that retrieval is a cue-dependent process that involves competition amongst co-activated memory traces. Recall that McGeoch’s response competition theory suggests that in an A-B, A-C paradigm both responses are available and the dominant response is expressed behaviourally. Hence, it was assumed that the MMFR would eliminate response competition because participants had unlimited time to produce both responses, and this would not be a problem if both target memory traces had been successfully retrieved (J. M. Barnes & Underwood, 1959). However, SAM suggests that traces compete during *retrieval* to the extent that some may be rendered temporarily inaccessible. In other words, both theories assign a role to competition, but SAM localises it within the retrieval process (Figure 1.1, red star), whereas McGeoch's theory suggests it is a conversion phenomenon (Figure 1.1, red square). More specifically, in SAM the probability of successfully retrieving a target memory trace is determined by its associative strength with all available retrieval cues (as in encoding-retrieval match, Tulving, 1983b), relative to competing traces that are also associated with the same cues (as in cue-overload theory, M. J. Watkins & Watkins, 1976). Note that there are strong similarities with the memory-as-discrimination account here, but with a clearer emphasis on the role of competition during retrieval rather than conversion. Because the potential interference occurs before all possible responses become available in memory, allowing the participant to provide multiple responses on an MMFR test does not alleviate the competition. According to SAM, the outcome of the Barnes and Underwood (1959) study (see Figure 1.5) occurs because the associative strength between cues and List 2 targets increases with the degree of List 2 repetitions. This simultaneously reduces the probability of List 1 target retrieval and increases the probability of List 2 target retrieval. As such, SAM can explain key interference phenomenon without recourse to storage-based impairment mechanisms such as unlearning or overwriting (Mensink & Raaijmakers, 1988).

1.2.7 Consolidation

We will now return to pick up the arc taken by biologically-oriented researchers following the inception of consolidation (Müller & Pilzecker, 1900). Consolidation is important from the perspective of memory updating because it implies that following a brief stablisation period, memory traces acquire a robust immunity to interference. This suggests that it would not be possible to simply overwrite old memory traces with new information. Although rejected by many psychologists, consolidation gained considerable traction within the neuroscience community (Glickman, 1961; McGaugh, 1966, 2000). These researchers were impressed by several lines of biologically-oriented evidence that appeared to offer converging support for the concept. We will outline the key aspects of this evidence first, before considering whether they are really as impressive as they might first seem.

1.2.7.1 Brain injury

Recall that the contrast between Short-Delay Interference and Long-Delay Interference conditions in the experiments of Müller and Pilzecker (1900) was taken to indicate that newly acquired traces experience a time-dependent decline in vulnerability to interference (Figure 1.2). It was quickly recognised that this temporally graded pattern was highly similar to clinical observations of patients who had suffered retrograde amnesia induced by brain injury (Burnham, 1903; McDougall, 1901; also see Russell & Nathan, 1946). For example, in a series of clinical case reports pre-dating the experiments of Müller and Pilzecker, the French psychiatrist Ribot documented that following brain injury, individuals tended to have poorer memory of the recent past relative to the distant past (Ribot, 1881, 1883). Ribot proposed "la loi de regression" (commonly know as "Ribot's Law"), which holds that older memory traces are more resistant to traumatic injury relative to more recently acquired memory traces. Although Ribot himself remained agnostic as to the locus of the forgetting[8], contemporary researchers tend to favour the interpretation that trauma-induced amnesia reflects disruption of a time-dependent consolidation process (A. S. Brown, 2002; Squire, 1992).

1.2.7.2 Electroconvulsive shock

Intensive investigations began in non-human animal laboratories to delineate the biological substrates of the consolidation process (for review see R. R. Miller & Springer, 1973; McGaugh, 1966; McGaugh & Dawson, 1971). Working with non-human subjects allowed for greater experimental control relative to clinical case studies, and enabled the deliberate use of invasive interventions that could directly influence the putative biological substrates of the consolidation process. When forgetting effects are observed under such circumstances they are broadly referred to as 'experimentally-induced amnesia'.

[8] Ribot (1883, p. 475) stated that "Two suppositions are equally warranted, viz., that either the registration of the prior states has been effaced; or that the retention of the anterior states persisting, their aptitude for being revived by associations with the present is destroyed. We are not in a position to decide between these two hypotheses."

In a prominent early study using electroconvulsive shock (ECS), Duncan (1949) first placed rats into an environment consisting of a 'dangerous' compartment and a 'safe' compartment. After a 10s period, a floor grid in the dangerous compartment delivered a mild electric foot shock, causing the rats to run into the safe compartment (an 'active avoidance' task). One of these learning trials occurred each day for a period of 18 days. The number of 'anticipatory runs' - when the rats ran into the safe compartment before the floor shock was delivered - was taken to indicate that the rats had learned a simple compartment-shock contingency. In the experimental groups, rats were subjected to a severe cerebral ECS after each trial. This was intended to perturb any organised physiological activity related to memory consolidation. Different groups of rats received the ECS at different time delays following learning (20s, 40s, 60s, 4m, 15m, 1h, 4h, and 14h). A control group received no ECS. Consistent with the notion of a time-dependent consolidation process, Duncan observed a temporally graded pattern of retroactive interference (Figure 1.7). The number of anticipatory runs, which was very low in the 20s-delay ECS group, gradually improved with an increase in the delay, and eventually plateaued in the 1h, 4h, and 14h groups, who had comparable performance to the no-ECS control group. Duncan concluded that ECS disrupted the progressive consolidation of information acquired during training, echoing the earlier findings of Müller and Pilzecker (1900). The time-dependent effectiveness of the ECS appeared to indicate that the consolidation process resolved somewhere in the period between 15m and 1h post-acquisition.

Further impetus for ECS research on consolidation came with the publication of an influential paper by Donald Hebb in the same year as Duncan's experiments were reported (Hebb, 1949). Hebb outlined a dual-trace theory in which newly acquired information was initially represented by transient reverberating activity in local neural circuits (short-term memory; STM). If left undisturbed for a period of time, this activity would lead to the structural changes associated with long-term memory (LTM). Hebb did not explicitly discuss the concept of consolidation, or cite Müller and Pilzecker (1900). Nevertheless, the concept of consolidation was analogous to the transfer of information from an STM to LTM format, and the two ideas became relatively synonymous (see McGaugh & Dawson, 1971). ECS was an especially useful tool for testing the theory because it could temporarily induce severe

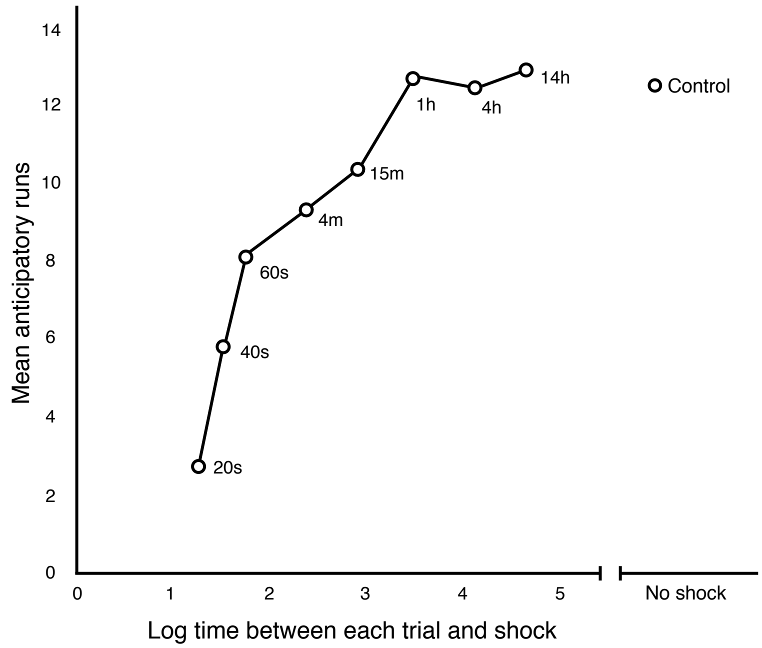


Figure 1.7: Findings of Duncan (1949). Mean anticipatory runs across 18 training trials as a function of trial-ECS delay. Adapted from Duncan (1949).

disruption of any organised brain activity, which, according to Hebb, must be allowed to persevere if the memory trace is to persist.

1.2.7.3 Protein-synthesis inhibitors

Hebb's theory also led researchers to explore the structural changes that might be necessary for the persistence of information as LTM. A popular approach involved the systemic infusion of pharmacological protein-synthesis inhibitors (PSIs; for review see Barraco & Stettner, 1976; Davis & Squire, 1984), such as anisomycin (ANI), and cycloheximide (CXM).

From amongst hundreds of studies, several key findings appear to have emerged (Davis & Squire, 1984), two of which are depicted for an illustrative case (Squire & Barondes, 1972) in Figure 1.8. Firstly, cerebral protein synthesis can be disrupted to a high degree (90%+) without having any major influence on ongoing behaviour (note the similarity of the learning curves for saline rats and CXM rats during the training stage). Secondly, during the test stage, performance in the CXM group was initially intact or only mildly impaired relative to the saline control group, but became progressively worse over time.

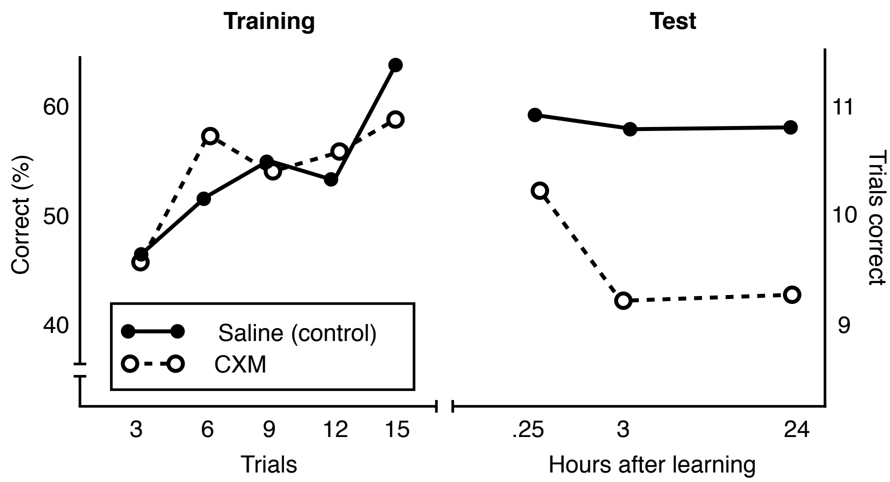


Figure 1.8: Findings of Squire and Barondes (1972). See main text for details. Adapted from Davis and Squire (1984).

In summary, several decades of biologically-oriented investigations across multiple species, paradigms, and intervention types converged on the same basic outcomes. Post-acquisition interventions (ECS, PSIs; also hypothermia, hypoxia, convulsant drugs, and brain lesions, for review see McGaugh & Dawson, 1971) cause amnesia that decreases as the time interval between training and intervention increases, to the point that they eventually become ineffective. The prevailing theoretical interpretation of these findings suggests that they reflect the progressive consolidation of memory traces into a fixed long-term memory format (Glickman, 1961; McGaugh, 1966, 2000).

1.2.7.4 Non-specific retroactive interference

As outlined earlier in this chapter, there was increasing disillusionment with the consolidation hypothesis amongst many psychologists as findings that appeared to conflict with those of Müller and Pilzecker (1900) began to accumulate. Retroactive interference effects did not appear to follow a temporally graded decline (e.g., J. A. McGeoch, 1933a), and varied as a function of the similarity of the interpolated learning materials (e.g., J. A. McGeoch & McDonald, 1931). However, recently both conclusions have been re-evaluated (Dewar et al., 2007; Wixted, 2004; also see Skaggs, 1933). Wixted (2004) for example, makes a distinction between *specific* retroactive interference (S-RI) and *non-specific* retroactive interference (NS-RI). S-RI arises during retrieval when two targets compete for a shared cue (as in an A-B, A-C paradigm). NS-RI arises when mental activity that involves the formation of new memory traces disrupts a post-acquisition consolidation process (as in an A-B, C-D paradigm). By this distinction, the experiments of Müller and Pilzecker (1900) involved NS-RI, as the interpolated task required learning new pairs of nonsense syllables that bore no direct relation to the original learning (in the standard notation of a paired-associates paradigm this was an A-B, C-D design). Similarly, in another experiment (Experiment$_{\mathrm{MP}}$ 35), retroactive interference effects were obtained when the interpolated task involved learning images of landscapes. Wixted (2004) argues that NS-RI effects will follow the temporally-graded pattern predicted by consolidation theory (as in Müller & Pilzecker, 1900), whereas S-RI effects will not. Therefore, because subsequent studies focused on S-RI (e.g., using an A-B, A-C paradigm) they may have overlooked important evidence in favour of consolidation theory.

As support for this hypothesis, Wixted (2004) re-evaluated a number of early A-B, A-C studies in which the temporal position of the interpolated task was varied over three time points (see Figure 1.9, Archer & Underwood, 1951; Newton & Wickens, 1956; Postman & Alper, 1946; Sisson, 1939). We will refer to these conditions as Immediate, Delayed, and Pre-Test. He noted that although several studies did not report statistically significant differences across time-points, a consistent 'inverted U' shaped pattern could be discerned in all of them (Figure 1.9 shows representative data from Newton & Wickens, 1956). This pattern is suggestive of two discrete forgetting processes. Firstly, in the Immediate condition, the close proximity of List 2 learning to List 1 learning leads to disruption of the consolidation of A-B associations (i.e., storage-based impairment through NS-RI). Secondly, in the Pre-Test condition, the close proximity of List 2 learning to the final test leads to temporary retrieval-based impairment of List 1 associations (i.e., S-RI). The Delayed condition benefits from being minimally affected by either of these processes.

Wixted's two-factor theory appears to capture two key aspects of the extant evidence. However, it is clear that considerably more empirical effort is needed. Firstly, given that many studies of experimentally-induced amnesia in animals indicate a time-course for consolidation

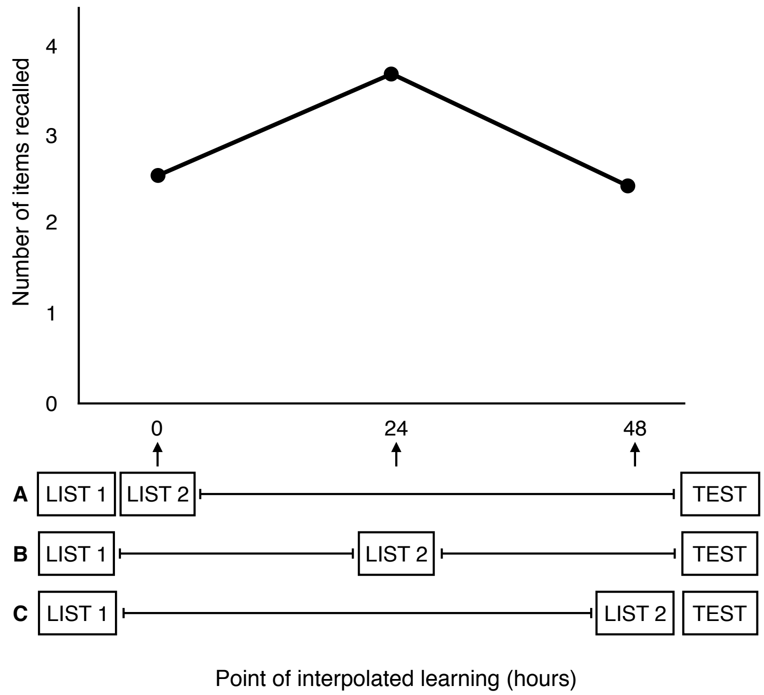


Figure 1.9: Design (bottom) and findings (top) of Newton and Wickens (1956). **A**: Immediate interference condition. **B**: Delayed interference condition. **C**: Pre-Test interference condition. Adapted from Wixted (2004).

in the manner of minutes (e.g., Duncan, 1949) or less (R. R. Miller & Springer, 1973), the reviewed experiments have poor temporal resolution. In other words, a more precise estimate of the temporal gradient needs to be obtained by systematically introducing new learning at short intervals post-acquisition. Secondly, as noted by Wixted (2004), the A-B, A-C design is not optimal for evaluating the theory because it conflates specific and non-specific interference effects. Thirdly, the robustness of the 'inverted-U' pattern is not firmly established because in some of the reviewed studies the effect of temporal interval was not statistically significant and sample sizes were typically rather small (e.g., Postman & Alper, 1946, n = 7 per condition; Archer & Underwood, 1951, n = 10 per condition). Limited reporting of statistical information and the absence of readily available raw data will also hinder any attempts at meta-analysis. Finally, it seems plausible that the performance 'dip' observed in the Immediate condition could be attributed to interruption of an active-rehearsal or retrieval-practice (self-testing) strategy rather than disruption of consolidation. In other words, performance in the Delayed condition could be higher because participants have had an opportunity to spontaneously practice the original learning material (even if that was not intended by the researchers). If such behaviour is occurring then it may not be obvious to the experimenter, and covert retrieval practice can be an effective method of improving subsequent recall (M. A. Smith, Roediger, & Karpicke, 2013).

Another explanation for poorer performance in the Immediate Delay vs. Intermediate Delay conditions (Figure 1.9, A vs B) is based on the concept of temporal distinctiveness (Lewandowsky, Ecker, Farrell, & Brown, 2012a; also see Sisson, 1939), formalised in the scale-independent memory, perception, and learning (SIMPLE) model (Brown, Neath, & Chater, 2007). The idea is that the retrievability of memory traces is a function of the degree to which they are 'isolated' along a relevant in psychological dimension such as time (similar to the notion of 'temporal context', see Figure 1.1, orange triangle). When traces are encoded in close temporal proximity, and further in the past, they are less distinctive, and consequently less retrievable. Critically, the model proposes that *relative* rather than *absolute* time is important: the degree of interference will depend upon the ratio of temporal distances for competing traces. For example, consider a memory trace 'X' that was encoded 20s ago. Retrievability of Trace X will be higher when a competing trace 'Y' was encoded 10s ago (ratio 2:1) compared to when Y was encoded 15s ago (ratio 4:3, example from Brown et al., 2007). According to SIMPLE, the temporally graded pattern of interference depicted in Figure 1.9 (A vs B), occurs because in the Immediate condition the two word lists are encoded in closer temporal proximity relative to the Delayed condition. Consequently, the lists are less discriminable, and retrievable, at the time of test (Lewandowsky et al., 2012a). In summary, the temporally-graded pattern of retroactive interference attributed to disruption of consolidation by Wixted (2004) can be accounted for by the SIMPLE model without recourse to any form of storage-based memory impairment.

An additional issue with Wixted's theory is the mechanism by which non-specific interference might disrupt the stabilisation process. Wixted (2004) suggests that "new learning draws on a limited pool of resources that may have otherwise been available to consolidate the original learning" (p.247). However, a recent experiment suggests that retroactive interference occurs even when no intentional learning is required (Dewar et al., 2007). Dewar et al.

first asked participants to learn 15 nouns. Following this, each participant completed one of six post-acquisition tasks ($n = 24$ per condition) that either did not involve intentional learning (spot-the-difference, mathematics, tone-detection) or involved an explicit instruction to memorise material (radio show, visual images). The sixth condition, in which participants were told to 'rest' and 'try not to think of the presented wordlists' (p.9) acted as the control condition. Performance was lower in all of the interpolated tasks relative to control and did not differ as a function of intentional learning. This demonstrates that retroactive interference can be caused by a task that does not require participants to form new memories, and thus would presumably have limited impact on the hypothetical pool of resources available for consolidation. Nevertheless, it is unclear to what extent the intentional learning manipulation actually suppressed processes involved in memory formation. It seems unlikely that participants would completely stop encoding information just because there is no intentional learning component to the task.

1.2.7.5 Multiple temporal gradients

Investigations of experimentally-induced amnesia appeared to place consolidation theory on a firm footing. However, a number of problematic findings started to emerge (Lewis & Maher, 1965). Firstly, although the findings of early studies (Duncan, 1949; Müller & Pilzecker, 1900) appeared to converge on an approximate time-frame (< 1h) in which consolidation processes reached completion, estimates diverged considerably in subsequent studies (for review see McGaugh, 2000), ranging from milliseconds (R. R. Miller, 1970) through to years (Squire, Slater, & Chace, 1975). Clinical reports also observed retrograde gradients extending over long delays, even decades (A. S. Brown, 2002; Russell & Nathan, 1946). This was difficult to reconcile, as it seemed highly unlikely that information would be maintained in an active (i.e., electrophysiological) state for such a long period of time (Crowder, 1989). In order to accommodate these findings, theorists drew a distinction between 'fast' *cellular* consolidation processes and 'slow' *systems* consolidation processes (Dudai, 2004; McGaugh, 2000). The 'cellular' process is thought to capture the transition of transient electrophysiological activity (STM) into structural synaptic changes that enable information storage (LTM), and is dependent on protein-synthesis. The standard account of the 'systems' process suggests that newly acquired information is temporarily stored in the hippocampus before being gradually distributed across neocortical sites. However, accounts vary considerably as to exactly what this 'reorganisation' involves, and whether information ever truly becomes hippocampal-independent (cf. McClelland, McNaughton, & O'Reilly, 1995; Nadel & Moscovitch, 1997; Squire & Alvarez, 1995; Winocur & Moscovitch, 2011).

1.2.7.6 Intact short-term memory

A consistent outcome of PSI studies was of intact performance on STM tests (i.e. shortly after the intervention) but impaired performance on LTM tests (see above, Davis & Squire, 1984). This finding can be accomodated by consolidation theory because the PSI presumably does not disrupt the electrophysiological activity associated with the STM trace, but disrupts the protein-synthesis dependent structural changes necessary for the trace to persist in the longer-term (McGaugh & Dawson, 1971). Therefore, intact performance on a short-term test simply shows that the STM is still intact. However, the same pattern has also been observed in ECS studies: performance is generally intact when tested shortly after ECS delivery (e.g., 15m delay) and impairments are only apparent on delayed tests (e.g., 60m delay; Geller & Jarvik, 1968b; McGaugh & Landfield, 1970; R. R. Miller & Springer, 1971). Because ECS likely eliminates the STM trace[9] before the putative structural changes that support LTM can be formed, it is not clear how this finding can be accommodated under standard consolidation theory. Instead, these findings suggest that the structural changes necessary to secure information in a latent state, and thus avoid obliteration via ECS, must already have taken place *prior to* ECS. In R. R. Miller and Springer (1971), ECS delivered just 10s after acquisition resulted in intact performance on a test taken 15m later, but impaired performance on a test taken 30m later. This indicates that (a) the processes (which could still be referred to as consolidation) necessary for formation of a structural memory trace occurred in less than 10s; and (b) the emergence of amnesia on a delayed test will require an

[9]ECS has a dramatic effect on extant neural activity triggering "massive neural firing followed by profound electrical silence throughout the brain" (R. R. Miller & Matzel, 2006, p. 492).

explanation other than disruption of consolidation.

1.2.7.7 Negative reinforcement

Invasive interventions are widely used because they are known to disrupt the putative substrates of the consolidation process. However, they also have a number of other side-effects that complicate interpretation. For example, when a rat is given an ECS, it seems plausible that it could act as a negative reinforcer (punishment). In the aforementioned study by Duncan (1949) for instance, in addition to learning that the 'dangerous compartment' was associated with a foot shock, rats in the experimental groups could also have learned to associate the 'safe compartment' with ECS. Naturally, this contingency would be more salient when the learning trial and ECS delivery occurred in close proximity, explaining the progressively higher performance levels with an increase in ECS delay. Duncan did attempt to address this point in a control experiment in which the rats received a shock (of equivalent intensity) to the hind legs in place of ECS. He reasoned that if the leg shock led to a similar pattern of retroactive interference as the ECS, then this would imply that ECS had no special influence on physiological processes occurring in the brain, and the negative reinforcement explanation would be more likely.

The findings of this control experiment were mixed. The 60s, 4m, and 45m delayed leg shock groups had similar performance to the no-shock control group, suggesting that the shocks were not acting as negative reinforcement. However, a 20s delayed leg shock group *did* show a substantial performance impairment (albeit not as pronounced as in the 20s delayed ECS group). Although Duncan does not give this finding much attention, it clearly leaves open the possibility that the delayed shock (ECS or foot shock) could be acting as a negative reinforcer. Although retroactive interference was not apparent in leg shock groups at longer delays, this could simply be because leg shock is not as unpleasant as the ECS. This seems plausible, given that ECS, as its name implies, typically induces the animal to convulse, and can even result in fatalities, as it did in this very study.

Indirect support for the negative reinforcement explanation can also be derived from studies that have used ECS whilst the experimental subjects are under anaesthetic (for review see

Lewis & Maher, 1965). The logic here is that when ECS is delivered under anaesthetic, the subject will not experience any punishing effects but any reverberatory activity should still be disrupted. In a study by Porter and Stone (1947) for example, rats completed a maze learning task and were then subjected to ECS either under anaesthetic (A+ condition) or unanaesthetised (A- condition). Larger performance impairments were observed in the A- condition compared to the A+ condition suggesting that the ECS could be acting as a negative reinforcer. However, as there was no condition where ECS was not applied, it is unclear if the ECS had any effect in the A+ condition. Therefore, it was not possible to rule out some effect of ECS on a consolidation process.

1.2.7.8 Recovery from experimentally-induced amnesia

In an earlier section (subsection 1.2.3) we outlined a number of recovery effects obtained under various conditioning protocols. These findings have generally been taken to provide compelling evidence against a storage-based interpretation of extinction and counter-conditioning (Bouton, 2002; Bouton & Moody, 2004). Below we examine a number of similar recovery effects that have been observed in studies where amnesia is induced through brain injury or invasive interventions such as ECS and PSIs (for review see Lewis & Maher, 1965; R. R. Miller & Springer, 1973; R. R. Miller & Matzel, 2006; Riccio & Richardson, 1984; Spear, 1973).

1.2.7.8.1 Spontaneous recovery

Clinical case studies showing that brain injury can lead to temporally graded amnesia have been interpreted as important evidential support for consolidation theory (A. S. Brown, 2002; Squire & Alvarez, 1995). However, it is also common to observe that the ‘forgotten’ time period occurring prior to injury ‘shrinks’ with the passage of time (e.g., Russell & Nathan, 1946; also see Kritchevsky, Zouzounis, & Squire, 1997). This is similar to ‘spontaneous recovery’ effects reported in extinction, counter-conditioning (Bouton, 2002), and paired-associate learning studies (A. S. Brown, 1976). Often recovery is almost complete, and only a brief residual amnesia remains (e.g., 30 minutes pre-injury). Similar findings have been obtained in clinical studies using ECT. For example, in a study by Squire et al. (1975; also see Squire, Slater, & Miller, 1981), patients who were receiving electroconvulsive treatment in an effort to relieve depression were asked before and after treatment to recall the names of various popular television programmes broadcast in preceding decades. A temporally-graded amnesia was observed, stretching back to around 3 years prior to treatment. However, recall performance had recovered considerably within 1-2 weeks following treatment, suggesting that the effect was only temporary. Together these findings indicate that a substantial amount of the amnesia observed in clinical studies cannot be attributed to storage deficits, as expected by consolidation theory.

Clinical observations can be difficult to interpret because of potential confounds, such as patients 'filling the gaps' in their memory with information from external sources. But spontaneous recovery has also been reported in more controlled non-human animal studies involving ECS-induced amnesia (Cooper & Koppenaal, 1964; Young & Galluscio, 1971; Zinkin & Miller, 1967). For example, Zinkin and Miller (1967) found that ECS delivered to rats after one-trial avoidance learning resulted in marked performance impairments after a 24h delay, but substantial recovery was observed on tests at 48h and 72h delays. However, this effect was not replicated in either of two subsequent attempts (Herz & Peeke, 1968; Luttges & McGaugh, 1967) and may have been driven by the use of repeated tests which provided opportunity for relearning (delayed tests were conducted within- rather than between-subjects). Nevertheless, a number of studies unaffected by this potential confound have also reported substantial spontaneous recovery when using PSIs as an intervention (e.g., Quartermain, McEwen, & Azmitia, 1972; Serota, 1971; Squire & Barondes, 1972; for review see Davis & Squire, 1984).

It would seem that spontaneous recovery effects in investigations of experimentally-induced amnesia parallel the situation with the paired-associates learning paradigm used in human studies (see above, A. S. Brown, 1976). The effects have been observed a sufficient number of times to consider them genuine phenomena, however it is clear that the parameters under which they do and do not occur are not well understood. It could be that the 'spontaneous' in spontaneous recovery merely reflects the fact that the relevant retrieval cue has not yet been identified. As discussed earlier, the effect could be a specific example of 'contextual renewal' where the temporal context guides the animal as to which behaviour is most appropriate

(Bouton, 1993). Alternatively, more rudimentary cues such as the presence of the same experimenter during the training and testing phases could be sufficient to aid recovery. Recovery could also be dependent on a number of factors which are generally not under experimental control, such as whether the animal attends to and encodes rudimentary cues during training, recognises the same cues during testing, and successfully retrieves the relevant representations from its memory store.

1.2.7.8.2 Reinstatement and renewal

As shown in extinction and counter-conditioning studies (Bouton, 2002), a reinstatement protocol in which the subject is presented with the US as a 'reminder' treatment prior to the test session can often lead to recovery from experimentally-induced amnesia. This is the case for both ECS (e.g., Geller & Jarvik, 1968a; Koppenaal, Jagoda, & Cruce, 1967; R. R. Miller & Springer, 1972) and PSIs (e.g., Quartermain, McEwen, & Azmitia, 1970; Radyushkin & Anokhin, 1999; Squire & Barondes, 1972). Other aspects of the training situation have also proven to be effective reminders, such as the CS (e.g., W. C. Gordon & Mowrer, 1980) and even the amnesic treatment itself (Hinderliter, Webster, & Riccio, 1975; Thompson & Neely, 1970). Re-exposure to the training context can also facilitate recovery (e.g., Sara, 1973) in a similar manner to 'renewal' effects following extinction training (Bouton, 2002).

Some researchers have argued that recovery effects do not necessarily rule out storage-based impairment effects (Cherkin, 1972; Gold & King, 1974; Squire, 2006). It could be that experimentally-induced amnesia is not an 'all-or-none' phenomenon and the reminder treatment provides a learning opportunity that summates with a residual memory trace (also see Rescorla & Heth, 1975). Although this may be plausible when a US-only reminder is employed, it is less clear how a CS-only reminder (e.g., Mowrer & Gordon, 1983) or context-only reminder (e.g., Sara, 1973) would work in this way. Indeed, presentation of an unpaired CS is equivalent to an extinction trial, which would normally lead to performance impairment rather than performance enhancement. Moreover, ECS-induced amnesia can be alleviated via pre-test delivery of psychostimulant drugs such as strychnine (W. C. Gordon & Spear, 1973; Sara & Remacle, 1977) and amphetamine (Quartermain, Judge, & Jung, 1988), which are unrelated to the CS-US contingency, and therefore cannot be said to summate with any residual information retained in memory storage.

1.2.7.8.3 Cue-dependent amnesia

A finding of central importance to consolidation theory is the consistent observation that the effectiveness of an amnesic agent decreases with the passage of time (McGaugh, 1966, 2000). Once the agent ceases to have an impact (i.e., performance is comparable to a no-intervention control), the memory trace can be considered robust to any subsequent intervention by that same agent. However, findings started to emerge that did not fit with this pattern. Misanin, Miller, and Lewis (1968) first replicated the typical finding of temporally graded amnesia in a tone-shock fear-conditioning paradigm: immediate post-acquisition ECS led to amnesia and ECS delivered after a 24h delay did not. However, when a retrieval cue (the CS) was provided 24h after training, immediately prior to ECS, a profound amnesia occurred. The effect could not be attributed to extinction induced by the CS reminder because a reactivation without intervention group showed only a small decline in performance.

The finding that a retrieval cue could 'reactivate' a presumably consolidated memory trace and render it vulnerable to an amnesic agent once again came to be known as 'cue-dependent amnesia' (Lewis, Bregman, & Mahan, 1972; Lewis, 1979). Cue-dependent amnesia was reported in several subsequent studies across a range of paradigms and intervention types (e.g., Bregman, Nicholas, & Lewis, 1976; DeVietti & Holliday, 1972; Gerson & Hendersen, 1978; Judge & Quartermain, 1982; Lewis et al., 1972; Mactutus, Ferek, George, & Riccio, 1982; A. M. Schneider & Sherman, 1968). Various aspects of the training situation proved to be effective cues, including the US, and CS-US re-pairing (Richardson, Riccio, & Mowrey, 1982). Notably, cue-dependent amnesia also appeared to be temporally graded (e.g., Judge & Quartermain, 1982; Mactutus et al., 1982), as with the post-acquisition amnesic effects associated with memory consolidation (Glickman, 1961; McGaugh, 1966).

This important finding suggested that the unidirectional transition from STM to LTM outlined under consolidation theory was not accurate. Lewis (1979) argued that it was the activity state of a memory trace, rather than the time since acquisition that determined its susceptibility to amnesic effects. He proposed a distinction between 'active memory' and 'inactive memory': at any given time, most memory traces will be in an inactive state, and have no effect on behaviour, however, various environmental cues will cause a subset of traces to transition to an active state. During this active state, traces can influence behaviour, but are also more vulnerable to interference. Over time, traces revert from an active to inactive state, explaining why the effectiveness of amnesic interventions is time-dependent (Judge & Quartermain, 1982; Mactutus et al., 1982). This is partly analogous to the distinction between memory traces and ecphoric memory outlined in Figure 1.1. Latent memory traces are equivalent to 'inactive memory'. Ecphoric memory refers to the state of activity created by the synergistic interaction of multiple retrieved memory traces ('active memory') and environmental input.

However, Lewis's theory failed to gain any considerable traction. A number of research teams had difficulty replicating cue-dependent amnesia effects (Banker, Hunt, & Pagano, 1969; Dawson & McGaugh, 1969; Gold & King, 1972; Jamieson & Albert, 1970; Squire, Slater, & Chace, 1976) and, unable to identify a reason for the discrepancy in findings, these authors rallied around consolidation theory. Consequently, the initial reports of cue-dependent amnesia had relative little impact on the prevailing zeitgeist (Riccio & Millin, 2003; Sara & Hars, 2006; Squire, 2006). The storage-retrieval debate also seemed unresolvable with the available empirical techniques and technologies, and entered into a period of prolonged inertia (Gold & King, 1974; Gold, 2006; R. R. Miller & Springer, 1973; R. R. Miller & Matzel, 2006).

1.2.8 Reconsolidation

Sporadic reports of cue-dependent amnesia continued to appear but were largely ignored (e.g., Przybyslawski & Sara, 1997; W. A. Rodriguez, Rodriguez, Phillips, & Martinez, 1993; Roullet & Sara, 1998; for review see Sara, 2000a). W. A. Rodriguez et al. (1993) and Przybyslawski and Sara (1997) attempted to accommodate the findings within the consolidation framework by proposing that reactivated memory traces re-enter a state of transient instability, from

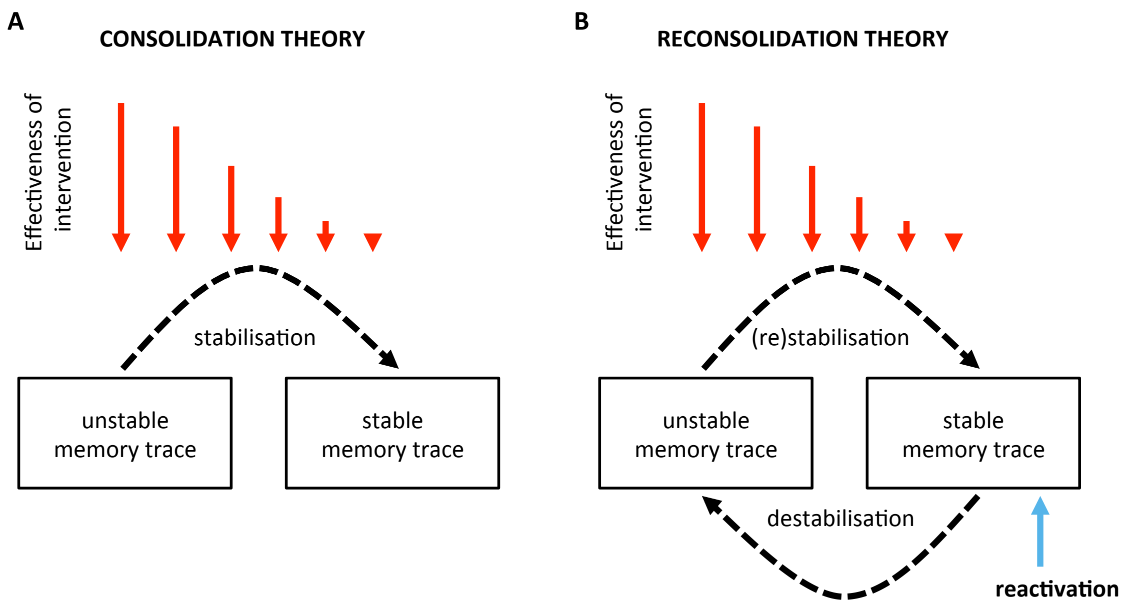


Figure 1.10: Schematic comparing consolidation theory and reconsolidation theory. Red arrows illustrate an idealised pattern of temporally-graded amnesia (the height of the arrows represents the effectiveness of a post-acquisition or post-reactivation intervention). **A**: Consolidation theory. **B**: Reconsolidation theory.

which they need to restabilise or 'reconsolidate'[10]. However, it was not until the publication of a study by Nader, Schafe, and Le Doux (2000a) that the field began to take the phenomenon seriously (Dudai, 2000; López, 2000; Nadel & Land, 2000; Nader et al., 2000b; Sara, 2000b). Perhaps cue-dependent amnesia effects seemed more appealing at this time because they were not being presented as a direct challenge to consolidation theory (as in Lewis & Maher, 1965; R. R. Miller & Springer, 1973; Riccio & Richardson, 1984), but merely required that it be partly adjusted and extended under the rubric of *re*consolidation theory (see Figure 1.10). The core features of consolidation theory were retained[11]: a newly acquired memory trace is

[10]The phrase 'reconsolidate' appears to have first been used by Spear (1973, p. 188), almost in passing, in a paper that critiqued consolidation theory, specifically its inability to account for cue-dependent amnesia and recovery effects. However, the idea was not given any clear specification until the later work of Sara and colleagues (Przybyslawski & Sara, 1997; Sara, 2000a).

[11]Note that most authors use reconsolidation to refer to a cellular-level restabilisation processes (e.g., Nader & Hardt, 2009) that bears similarities to cellular consolidation rather than systems consolidation (Dudai, 2004). However, several authors have considered a potential role for systems-level reconsolidation

initially unstable and progressively stabilises ('consolidates') over time, eventually becoming resistant to amnesic agents (McGaugh, 1966, 2000). However, according to reconsolidation theory a previously consolidated memory trace can be destabilised via memory retrieval, rendering it once again vulnerable to interference until is has restabilised ("reconsolidated", see Nader, 2003a; Nader & Hardt, 2009).

The study by Nader et al. (2000a) was especially compelling because it involved direct infusion of a PSI into sites assumed to eventually harbour the target memory traces (previous studies tended to involve systemic infusion, see Davis & Squire, 1984), and included a number of critical control groups. The study built on Schafe and LeDoux (2000) who reported evidence for consolidation in an auditory fear conditioning paradigm. In that study, rats were trained to associate a tone (CS) with a foot-shock (US). Learning was indexed by the proportion of time animals exhibited a fear response ('freezing') following presentation of the tone in isolation. Infusion of the PSI anisomycin into the lateral and basal amygdala (LBA) shortly after training, led to the typical finding that amnesia was absent on a short-term test (STM), and present on a delayed test (LTM). Additionally, performance was intact if the PSI was delivered 6h after training, indicating that the effectiveness of the intervention was time-dependent. These findings implied that the memory trace representing the CS-US association was consolidated within 6h, and should therefore be resistant to subsequent interference.

However in Nader et al. (2000a), rats who were exposed to the tone (CS) 24h after training, and then received the PSI infusion, also showed amnesia at test, echoing previous findings of cue-dependent amnesia (see paragraph 1.2.7.8.3). Critically, the amnesia was dependent on both the intervention (see Figure 1.11, panel A), and the reactivation (see Figure 1.11, panel B) components of the procedure. Furthermore, as in the Schafe and LeDoux (2000) consolidation study, the effectiveness of the intervention was time-dependent: amnesia was substantially reduced when anisomycin was infused after a 6h delay (see Figure 1.11, panel C). Similarly,

(e.g., Debiec, LeDoux, & Nader, 2002; McKenzie & Eichenbaum, 2011). Our focus will be on the cellular-level interpretation as it is the dominant rationale for empirical investigations in both the non-human animal (Nader & Hardt, 2009) and human (Schiller & Phelps, 2011) literatures.

performance was intact on a STM test (4h after reactivation), but amnesia was observed on a LTM test (24h after reactivation; see Figure 1.11, panel D). In summary, the emergence of an amnesic effect on a delayed test was contingent on the time-dependent interaction of the reactivation and intervention, consistent with the key tenets of reconsolidation theory (Nader et al., 2000b; Nader, 2003a; Nader & Hardt, 2009), and previous demonstrations of cue-dependent amnesia (e.g., Lewis et al., 1972; Mactutus et al., 1982; Misanin et al., 1968).

Over the subsequent decade and a half, and up to the present time of writing, there has been a rapid proliferation of empirical and theoretical reports related to memory reconsolidation (see Figure 1.12, for general reviews see Alberini, 2011; Besnard, Caboche, & Laroche, 2012; Dudai, 2004, 2012; Nader & Hardt, 2009; Nader, 2015; Tronson & Taylor, 2007). Reconsolidation effects have been reported across a wide range of species (e.g., rats, Nader et al., 2000a; crabs, Pedreira, Pérez-Cuesta, & Maldonado, 2004; honeybees, Stollhoff, 2005; medaka fish, Eisenberg, Kobilo, Berman, & Dudai, 2003; humans, Hupbach et al., 2007), experimental paradigms (e.g., auditory fear conditioning, Nader et al., 2000a; spatial water maze, Morris et al., 2006; inhibitory avoidance, Milekic & Alberini, 2002; list-learning, Hupbach et al., 2007; motor-sequence learning, Walker, Brakefield, Hobson, & Stickgold, 2003), and amnesic agents (PSIs, Nader et al., 2000a; RNA synthesis inhibition, Sangha, Scheibenstock, & Lukowiak, 2003; anaesthesia, Eisenberg et al., 2003; new learning, Hupbach et al., 2007). Based on this large body of evidence, it has been proposed that reconsolidation represents a fundamental memory process (J. L. Lee, 2009; Nader & Hardt, 2009).

1.2.8.1 Reconsolidation and memory updating

Reconsolidation theory suggests that there are two periods of transient plasticity during which a memory trace is open to modification: shortly after acquisition and shortly after reactivation. According to this idea, the accuracy and relevance of information stored in memory traces is routinely maintained by incorporating new environmental input during recurring cycles of reconsolidation that are triggered by memory retrieval. Thus, reconsolidation could be a process that enables both the persistence and plasticity of the memory trace.

This proposal may seem surprising when the majority of the evidence-base for reconsolidation consists of studies in which amnesia is induced through invasive pharmacological agents that an animal would be unlikely to encounter in its natural environment. Nevertheless, theorists argue that reconsolidation-mediated memory updating can explain the 'dynamic nature of memory', and place great emphasis on classic findings in the cognitive psychology literature, such as retroactive interference in the paired associates paradigm, and the misinformation effect (Dudai, 2009; Hardt et al., 2010; Hupbach et al., 2007; Schiller & Phelps, 2011). For example, in reference to the misinformation effect Hupbach et al. (2007, p. 51) state, "the idea that episodic memory traces can be modified by post-event information has long been accepted in cognitive psychology" (cf. subsection 1.2.5).

Critically, these authors are assuming that dynamic behaviour necessitates dynamic memory traces. This conflates behavioural plasticity (the capacity of an organism to modify its behaviour over time in response to changes in its environment) with physiological plasticity (the capacity of memory traces to be modified with new information). In fact, the general picture that emerges from the evidence reviewed in this chapter suggests that behavioural plasticity is not necessarily contingent on modification of existing memory traces. For instance, retroactive interference effects in the paired associates learning paradigm can largely (or even entirely) be attributed to retrieval dynamics rather than storage-based trace alterations (see subsection 1.2.6, Mensink & Raaijmakers, 1988). There is also compelling evidence that misinformation effects do not involve overwriting of event memory traces, but reflect the influence of non-memory related factors operating during conversion (see subsection 1.2.5 and Chapter 4, M. McCloskey & Zaragoza, 1985a). Similarly, behavioural interference paradigms in the non-human animal literature, such as extinction and counter-conditioning, do not appear to involve destructive updating of previously acquired memory traces (see subsection 1.2.3, Bouton, 2002).

Nevertheless, the reconsolidation-updating theory makes an important testable prediction: If a consolidated memory trace is retrieved prior to an incidence of new learning, then it becomes temporally vulnerable to overwriting. Although the use of new learning as a post-retrieval intervention is relatively rare in non-human animal studies (but see Monfils, Cowansage, Klann, & LeDoux, 2009; Richardson, Riccio, Jamis, Cabosky, & Skoczen, 1982), it is the typical approach employed in reconsolidation investigations with human participants (for review see Agren, 2014; Schiller & Phelps, 2011). This is partly because ethical limitations restrict the use of invasive amnesic agents in humans (although see Kindt, Soeter, & Vervliet, 2009; Kroes et al., 2014), but also because of the potential clinical benefits that reconsolidation-mediated memory updating appears to offer (R. D. Lane, Ryan, Nadel, & Greenberg, 2015; Schwabe, Nader, & Pruessner, 2014). This literature will be reviewed extensively in Chapter 2.

1.3 Overview of thesis

Over the course of a century a recurring debate has been borne out across multiple fields of scientific enquiry in an enduring search for the loci of forgetting (Figure 1.1). The outcome of this debate speaks to a broader question of how the memory system can allow stored information to persist whilst simultaneously maintaining a capacity for plasticity. Is it necessary that old information be destroyed or overwritten in order to accommodate updating and plasticity? Or can individuals learn new information whilst allowing existing memory traces to persist?

In this thesis, we will examine the claim that memory reconsolidation enables memory updating through a process of overwriting. Chapter 2 provides a general critique of reconsolidation theory and a comprehensive review of existing investigations of reconsolidation-mediated memory updating in humans. Chapter 3 outlines several attempts to replicate a key demonstration of this phenomenon (Experiments 1-7). Chapter 4 reports a study (Experiments 8-10) designed to evaluate whether a hybrid reconsolidation-misinformation paradigm will elicit memory impairment effects consistent with the overwriting hypothesis, or whether event memory traces continue to persist despite post-retrieval exposure to contradictory misinformation. Chapter 5 evaluates whether a reconsolidation-overwriting mechanism might facilitate correction of semantic misconceptions (Experiment 11). In Chapter 6, we will move away from reconsolidation and explore the consequences of memory persistence in the context of the continued influence effect (Experiment 12). Finally, Chapter 7 will provide a summary and integrative discussion of the findings presented in this thesis and suggest avenues for future research.

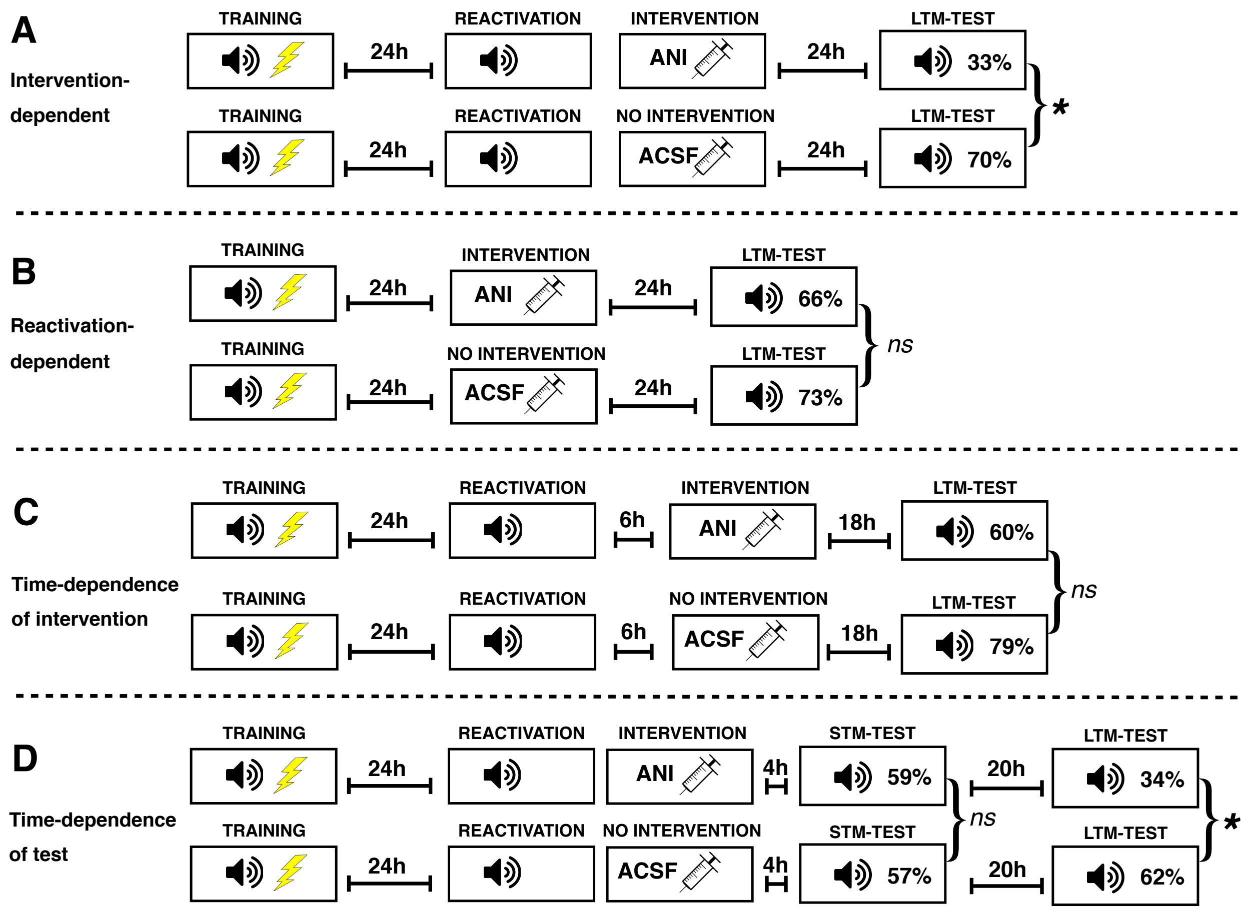


Figure 1.11: Design and findings of Nader et al. (2000a). *Training* consisted of a single presentation of the CS (tone, 30s) which coterminated with the US (foot shock, 1s). *Reactivation* consisted of a single presentation of the CS. *Interference* involved infusion of ANI (anisomycin) into the lateral and basal amygdala (LBA). *No Interference* involved infusion of ACSF (artificial cerebrospinal fluid, i.e., vehicle) into the LBA. *STM-Test* (short-term memory test) and *LTM-Test* (long-term memory test) consisted of three presentations of the CS. Percentages shown in Test boxes refer to the mean time spent immobile (freezing) during the first 30s CS trial of test. $*$ = statistically significant at the .05 level. *ns* = not statistically significant. The effect was **A**: Intervention dependent; **B**: Reactivation dependent; and indicated the **C**: Time-dependence of the intervention; and **D**: Time-dependence of the test.

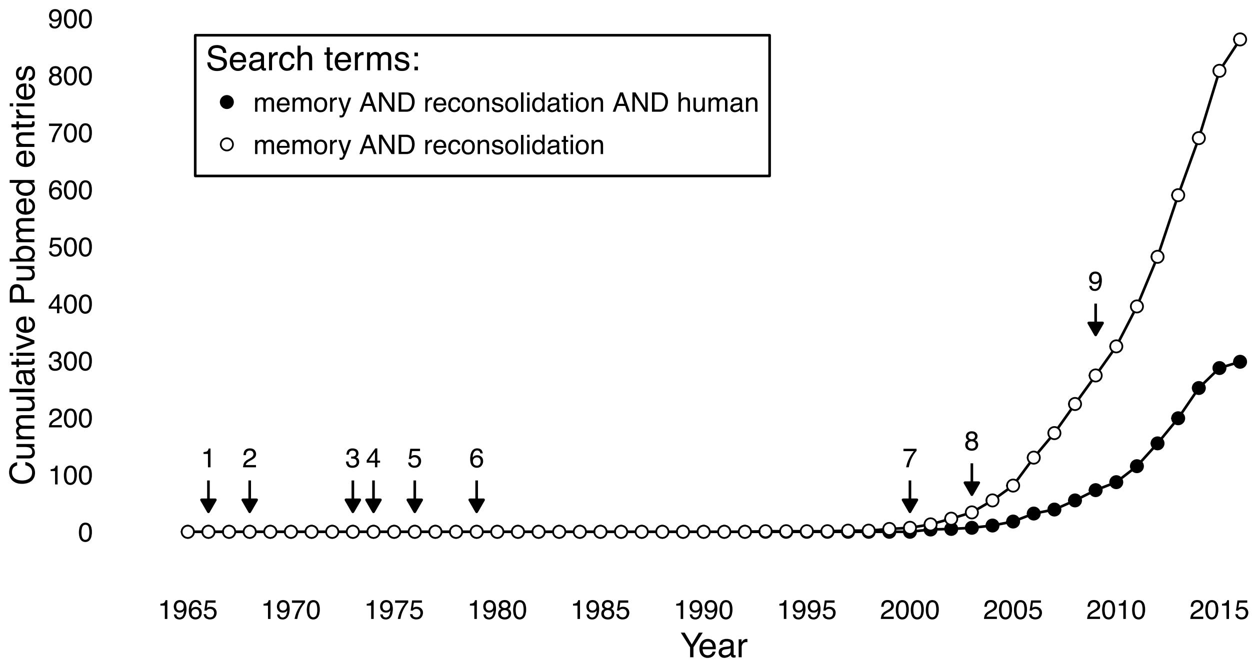


Figure 1.12: Research trends for memory reconsolidation. Cumulative Pubmed entries for the search terms [“memory” AND “reconsolidation”] and [“memory” AND “reconsolidation” AND “human”] (search conducted 23/07/2016). Note that these searches are only intended to provide an approximate insight into research trends on memory reconsolidation. The searches (a) overlap; and (b) include both empirical and non-empirical articles (e.g., reviews, opinion pieces). Arrows and numbers indicate key developments in the field of reconsolidation. **1**: McGaugh (1966) publishes foundational review paper on memory consolidation theory; **2**: Misanin et al. (1968) reports cue-dependent amnesia effect using ECS following memory reactivation in rats; **3**: Miller and Springer (1973) outline major critique of memory consolidation theory; **4**: Gold and King (1974) respond with a defense of consolidation theory; **5**: Squire (1976) is unable to obtain cue-dependent amnesia effect using ECS in humans; **6**: Lewis (1979) outlines an alternative to consolidation theory in which memory traces can shift bidirectionally between active and inactive states; **7**: Nader et al. (2000a) demonstrate cue-dependent amnesia using a PSI in rats. Their study has a profound impact upon the field; **8**: Walker et al. (2003) publish what is generally considered to be the first compelling demonstration of reconsolidation in humans; **9**: Nader and Hardt (2009) publish major review of reconsolidation theory and conclude that reconsolidation respresents a “fundamental mnemonic process”. Figure adapted and extended from Besnard et al. (2012).

Chapter 2

Does reconsolidation enable memory trace updating? A literature review.

> "We are faced with the impossible challenge of proving that the memory trace does not exist."
>
> — Sara and Hars (2006, p.518)

As outlined in Chapter 1, reconsolidation theory has gained considerable attention in recent years (see Figure 1.12), and numerous review articles suggest that it has received broad empirical support across a range of species, intervention types, and paradigms (Besnard et al., 2012; Nader & Hardt, 2009; Schwabe et al., 2014; Tronson & Taylor, 2007). The popularity of the theory likely stems from its ability to maintain the explanatory coverage of its forebear, consolidation theory, whilst also accommodating previously problematic demonstrations of cue-dependent amnesia (see Chapter 1, R. R. Miller & Springer, 1973; Riccio & Richardson, 1984). Consequently, reconsolidation theory preserves many of the core tenets of consolidation theory whilst addressing one of the principle evidentiary lines that appear to undermine its validity (see Figure 1.10, Lewis & Maher, 1965; R. R. Miller & Springer, 1973; R. R. Miller & Matzel, 2006; Riccio & Richardson, 1984; Riccio, Millin, & Bogart, 2006).

Similarly, the idea that 'reconsolidation is the mechanistic instantiation of memory updating' (J. L. Lee, 2009, p. 418), has received widespread attention (e.g., Besnard et al., 2012; Dudai & Eisenberg, 2004; Exton-McGuinness, Lee, & Reichelt, 2015; Hardt et al., 2010; Nader et al., 2000b; Sara, 2000b; Tronson & Taylor, 2007). According to this account, the reconsolidation process that is triggered upon trace retrieval opens a transient window of plasticity during which the trace can be updated with incoming environmental information. Investigation of such a theory would naturally require studies in which the post-retrieval intervention involves an episode of new learning. However, as indicated in Chapter 1, the most common intervention procedure entails delivery of invasive amnesic agents such as electroconvulsive shock (ECS) or protein-synthesis inhibitors (PSIs) to non-human animals

(Tronson & Taylor, 2007; Nader & Hardt, 2009). Although physiological interventions are intended to directly disrupt the putative molecular substrates of reconsolidation, considerable ambiguity surrounds the envisioned mechanism by which a behavioral intervention might influence these same processes. Indeed, reconsolidation-updating claims are typically cast in broad terms (Dudai, 2009; Exton-McGuinness et al., 2015; Hardt et al., 2010; J. L. Lee, 2009), and the qualitative nature of empirical outcomes taken to support these claims varies considerably. In some cases, the updating process is considered to be *constructive*, allowing for the 'incorporation' or 'integration' of new information 'into' existing memory traces (e.g., Forcato et al., 2007; Hupbach et al., 2007; St Jacques & Schacter, 2013). Conversely, in other cases the updating process is considered to be *destructive* and results in the content of memory traces being selectively 'rewritten' by post-retrieval new learning (J. C. K. Chan & LaPaglia, 2013; Schiller et al., 2010; Walker et al., 2003). Finally, when the reconsolidation process is left unperturbed, it is thought to *strengthen* the underlying trace (Finn & Roediger, 2011; Forcato et al., 2011; Sara, 2000b).

Although the consequences of post-retrieval new learning have rarely been explored in non-human animal studies (cf. Monfils et al., 2009), the use of behavioural interventions is much more common in the human reconsolidation literature. Due to ethical constraints, the use of invasive amnesic agents is limited (Schiller & Phelps, 2011). Some human studies have used less toxic pharmacological interventions, for example the beta-blocker propranolol (e.g., Kindt et al., 2009), or have been able to capitalize on opportunities where electroconvulsive shock treatment is being used as a treatment for patients with depression (Kroes et al., 2014). However, the vast majority of human investigations have used new learning as a post-retrieval intervention (for previous reviews see Agren, 2014; Schiller & Phelps, 2011; Schwabe et al., 2014). Unfortunately, limitations on the use of ECS or PSIs complicates direct translation of protocols and findings between the two literatures. This has resulted in considerable heterogeneity in the various methodological parameters employed across human reconsolidation investigations. This methodological plurality may afford greater external validity, but other important factors such as reliability and internal validity appear to have suffered. As a result, it has become increasingly difficult to ascertain whether there is any compelling evidence to suggest that (a) reconsolidation occurs in humans; and (b) reconsolidation enables memory traces to be updated (overwritten) with new information.

In this chapter, we will summarise and evaluate the evidentiary basis of these two proposals. If reconsolidation enables overwriting of memory traces in humans, it could have profound theoretical (Nader et al., 2000b), clinical (Schwabe et al., 2014), and ethical (Hui & Fisher, 2014) implications. For example, the ability to erase 'pathological' memory traces that contribute to post-traumatic stress disorder, addiction, and phobias, offers the potential of lasting relief from these conditions (Schwabe et al., 2014). More broadly, compelling evidence for reconsolidation-mediated memory updating would represent a significant new development in the debate outlined in Chapter 1 regarding the loci of forgetting. Further exploration of reconsolidation could help to elucidate how persistence and plasticity are simultaneously maintained in the human memory system.

In the first section (section 2.1), we outline 10 key criteria that should ideally be met if a study is to provide a robust test of reconsolidation theory. We then systematically and comprehensively evaluate the extent to which these criteria are met in human reconsolidation investigations that have employed behavioural interventions. In the second section (section 2.2), key case studies in the human reconsolidation literature will be described and discussed in detail in order to highlight some of the idiosyncratic issues that can arise in such investigations. The final section (section 2.3), will take a broader 'meta-theoretical' perspective on the concept of reconsolidation, and examine how the field has handled empirical successes and failures. The chapter concludes with some concrete recommendations that may foster a more fruitful approach in future investigations of reconsolidation.

2.1 Systematic review

2.1.1 Criteria

Decades of investigation into consolidation, cue-dependent amnesia, and subsequently reconsolidation, have established a number of key criteria that should be met if a study is to provide a robust test of the theory that a post-acquisition/post-retrieval (re)stabilisation process is necessary for the (re)storage of the memory trace. These criteria are especially important because, like its forebear consolidation, reconsolidation is a “curiously ‘invisible’ construct” (Lewandowsky et al., 2012a, p. 37) that must be indirectly inferred from behavioural data (Nader, 2006; Tronson & Taylor, 2007). Given the multifaceted and often opaque stages that intervene between an episode of new learning and subsequent learning-dependent performance (Chapter 1, Figure 1.1), it can be extremely difficult to design experiments that isolate the mechanisms operating during any single stage.

The criteria outlined here are often implicit in the control groups that researchers use in their investigations of reconsolidation (e.g., Nader et al., 2000a) and occasionally they are stated explicitly in review papers (e.g., Agren, 2014; Dudai, 2004; Tronson & Taylor, 2007; Schiller & Phelps, 2011). However, the overlap between articles is not always complete, and adherence to the criteria in empirical investigations varies considerably, especially in the human reconsolidation literature. Here we have attempted to collate and precisely define criteria that are implemented sporadically across the field and then systematically evaluate the extent to which they have been met in studies using behavioural inventions with human participants. A detailed table containing case-by-case outcomes of the review (Table 2.5) is provided at the end of the chapter. The table is also available as a .csv file (https://osf.io/42k3u/). A summary graph representing these outcomes is shown in Figure 2.2. Studies that have used non-behavioural (i.e., pharmacological or electrophysiological) interventions will not be discussed in great detail in the present review. The criteria should be considered minimum standards given the present state of the field, and it is likely that they will need to be adjusted and supplemented as the field develops.

In the remainder of this chapter we will use the following short-hand. When a criterion (e.g., Criterion 1, C1) is met, this will be indicated with a ‘Y’ symbol (e.g., $C1^{Y}$). In this case, the study has either successfully implemented a condition necessary for demonstrating a reconsolidation effect, or observed a result consistent with the predictions of reconsolidation theory. When a criterion has been implemented/assessed but not met, this will be indicated with a ‘N’ symbol (e.g., $C1^{N}$). In this case the study has either failed to meet the conditions necessary for a robust test of reconsolidation theory, or else it has observed a result inconsistent with the predictions of the theory. When a criterion referring solely to findings (as opposed to methodological conditions) has not been assessed, this will be indicated with a '?' symbol (e.g., $C1^?$). In this case, the study does not provide sufficient information for us to assess the particular aspect of the theory to which the criterion refers. In some cases, some idiosyncratic reason makes it unclear whether a criterion has been met or not. This will be indicated with a 'U' symbol (e.g., $C1^U$) and additional explanation will be provided either in the table footnotes (Table 2.5) or main text. Finally, in some cases criteria are not applicable. For example, if a study finds that performance is not contingent on provision of a reminder (i.e., $C4^N$), then it is no longer necessary to assess the 'Absence of recovery' criterion (C10), because there is no impairment from which to recover. These situations will be indicated in Table 2.5) with a ':'. We will also use a short-hand to refer to the key experimental groups that enable the criteria to be met. These group codes are displayed in Table 2.1.

Table 2.1: Group codes for key conditions in reconsolidation experiments.

Group	Group Code	Related Criterion
Reminder and intervention	R^+I^+	All
No reminder, intervention	R^-I^+	C4, C7
Reminder, no intervention	R^+I^-	C5, C7
No reminder, no intervention	R^-I^-	C6, C7
Reminder, delayed intervention	R^+I^D	C8
Reminder and intervention, short-term memory test	$R^+I^+T^{STM}$	C9
Reminder and intervention, recovery test	$R^+I^+T^{REC}$	C10

The criteria are shown in Table 2.2 alongside a canonical reconsolidation protocol in Figure 2.1. A brief introduction is provided here before each criterion is described in greater detail in the following sections. When an existing, consolidated (C1) memory trace is reactivated via a reminder (C4), it is destabilised and transiently open to modification by a variety of post-retrieval (C2) interventions (C5). Modification effects can be detectable through a change in trace-dependent performance between baseline (established during training, or by the reminder protocol itself, C3) and the Test Stage. Any modification effect will be dependent on the interaction of the reminder and intervention (C7) and cannot be solely attributed to the independent influences of the reminder and/or intervention (C6). Interventions will operate with decreasing effectiveness along a temporal gradient until progressive restabilisation processes have terminated and the trace has been reconsolidated (C8). Consequently, an appropriate intervention delivered within the time-window of restabilisation will serve to modify the existing memory trace in a manner that is permanent and shows limited propensity for recovery (C10). Because reconsolidation is thought to involve a time-dependent process, the consequences of the intervention should only become apparent on a delayed test (i.e., post-reconsolidation) and performance may be unaffected on an immediate (or short-delay) test (C9).

Table 2.2: Criteria for a compelling demonstration of reconsolidation.

Code	Criterion
C1	Trace has consolidated
C2	Reactivation precedes intervention
C3	Baseline-comparison
C4	Reactivation-dependent
C5	Intervention-dependent
C6	Control for independent effects
C7	Reminder*intervention interaction
C8	Time-dependence of intervention
C9	Time-dependence of test
C10	Absence of recovery

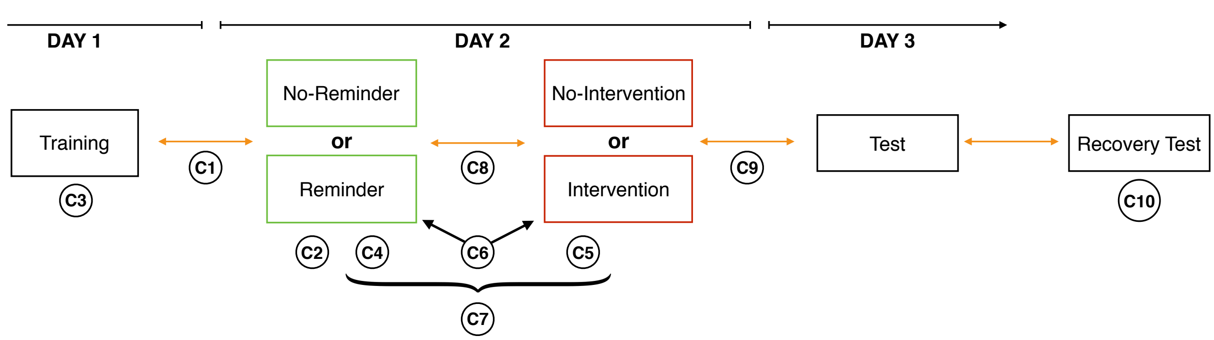


Figure 2.1: Canonical reconsolidation protocol and necessary criteria for a compelling demonstration of reconsolidation (see Table 2.2).

2.1.2 Literature search

Relevant studies were identified by searching the PsychInfo database for the keyword term "reconsolidation" with the search limiters "human", and "English language" (final search conducted 23rd June, 2016). This search returned 334 records. Each record was checked manually to assess whether it referred to an empirical assessment of memory reconsolidation theory (articles had to report novel data; review and opinion articles were not included). Finally, this collection of records was supplemented with any additional articles meeting the above parameters that were known to the present author. Only studies that employed behavioural interventions are included in the systematic aspect of this review. Note that the term 'cases' will be used to refer to individual experimental conditions in which a reconsolidation effect is expected to occur (i.e., R^+I^+ conditions, see Table 2.5). Cases are evaluated based on associated control conditions reported in the same study (i.e., research article). In total, 168 cases reported in 61 different studies were included in the systematic review.

2.1.3 Criterion 1: Trace has consolidated

2.1.3.1 Description of criterion

A key prediction of reconsolidation theory is that a previously consolidated memory trace can be destabilised by a reminder cue, thus initiating a restabilisation process. It is therefore necessary to first establish that the target memory trace has consolidated before it can be claimed that it has subsequently undergone reconsolidation[12] (Nader & Hardt, 2009; Tronson & Taylor, 2007). However, as discussed in subsection 1.2.7, there is no direct way to measure that a trace has consolidated. In the absence of definitive biological indicators of trace stabilisation, the operational definition of a consolidated trace is its immunity to post-acquisition interventions (McGaugh, 1966). However, because 'consolidation windows' appear

[12] One could object by disputing the relative merits of consolidation theory itself (see subsection 1.2.7, Lewis, 1979; Lewandowsky et al., 2012a) but this criterion is necessary for reconsolidation theory to at least be internally consistent.

to vary as a function of the intervention used (Dudai, 2004), immunity to one intervention does not necessarily imply immunity to all interventions. In other words, establishing that a trace has 'consolidated' in the operational sense, does not necessarily mean it has consolidated in the theoretical sense. At the molecular level, this could be taken to imply that different aspects of the consolidation process have different temporal trajectories (Dudai, 2004; Gold, 2006).

Naturally, if one cannot demonstrate that a trace has consolidated then it is difficult to argue for the necessity of reconsolidation. Indeed, some theorists have proposed that apparent reconsolidation effects might be more accurately characterised as instances of 'lingering consolidation' (Alberini, 2011; Dudai & Eisenberg, 2004). Nevertheless, if one can establish a time-window of effectiveness for a specific intervention (e.g., the intervention loses its efficacy after 6h, Schafe & LeDoux, 2000), and then demonstrate that reactivation opens a new time-window of effectiveness for that same intervention (Nader et al., 2000a), then many would consider this favourable evidence for the reconsolidation account (López, 2000; Nadel & Land, 2000; Nader et al., 2000b; Sara, 2000b).

In practice, few studies follow this approach and instead assume that the memory trace has consolidated after a 24h period has elapsed since acquisition (Dudai, 2004). This 'rule of thumb' can be traced back to early investigations of cue-dependent amnesia. For example, in a review of the literature, Miller and Springer (1973, p. 74) noted that "The time between training and reactivation is usually 24 hours, an interval sufficient to insure consolidation by most any standard and inadequate to permit appreciable forgetting." For the purposes of this review, we will also adopt this 24h standard as our criterion for a 'consolidated' memory trace. If a shorter time interval is employed on the basis of an operational demonstration that the consolidation window has closed, that would also satisfy the criterion. However, none of the reviewed studies took this approach. Time-intervals longer than 24h are also more than acceptable. Indeed, if routine memory updating occurs via reconsolidation (J. L. Lee, 2009) then traces older than 24h should also undergo reconsolidation. Memory age is currently a contested boundary condition on reconsolidation theory (see section 2.3).

Meeting this criterion ($C1^{Y}$) requires either that there is a minimum 24 interval between acquisition and reminder, or else the researchers have timed the reminder based on an operational demonstration that the consolidation window has closed. Otherwise, the criterion will not have been met ($C1^N$).

CRITERION 1: Trace has consolidated

At a minimum, reactivation should occur 24 hours after acquisition.

2.1.3.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C1), and summary data are presented in Figure 2.2. 148 out of 168 cases met the criterion (88%). The remaining 20 cases did not meet it (12%).

The studies that did not meet this criterion tended to involve all key stages (see Figure 2.1) being conducted within a single experimental session (e.g., Diekelmann et al., 2011; J. C. K. Chan & LaPaglia, 2013, $\text{Experiment}_{\text{CL}}$ 1, $\text{Experiment}_{\text{CL}}$ 4). Based on typical operational estimates of the consolidation window (see above), it seems unlikely that the target traces would have fully stabilised by the time the reminder was delivered in these studies. Nevertheless, without a direct measure of trace stablisation, there is no way to know for certain if the target traces had initially consolidated or not. In the vast majority of cases, there was a 24h or longer interval between the training and reminder stages, allowing reasonable confidence that the target memory traces had consolidated by the time they were reactivated.

2.1.4 Criterion 2: Reminder precedes intervention

2.1.4.1 Description of criterion

Because reactivation is necessary to trigger reconsolidation, the intervention must be delivered *after* the reminder if it is to disrupt the reconsolidation process. This may seem obvious,

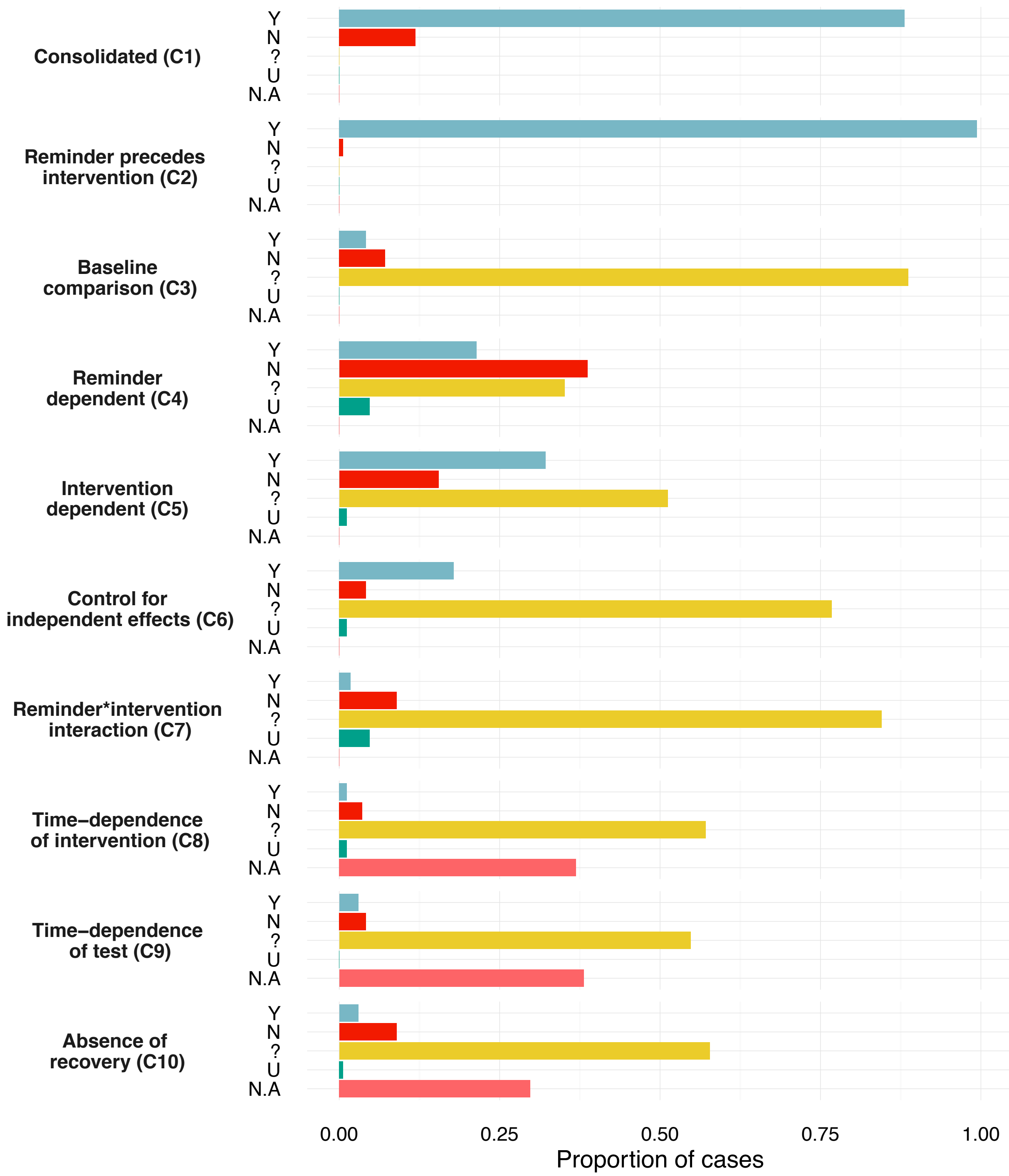


Figure 2.2: Adherence to reconsolidation criteria in 168 cases, determined independently across criteria. *Y* = Criterion met. *N* = Criterion not met. *?* = Criterion not implemented/assessed. *U* = Unclear if criterion met. *N.A* = Criterion not applicable.

but practical constraints sometimes make it necessary to deliver the intervention *before* or *during* the reminder (see below). This complicates interpretation of the observed effect because the intervention could have directly influenced retrieval, rather than the subsequent reconsolidation process (Nader & Hardt, 2009; Schiller & Phelps, 2011; Tronson & Taylor, 2007).

Interventions are often delivered *before* the reminder in studies using propranolol (which are not included in the systematic aspect of this review; e.g., Kindt et al., 2009; Schwabe, Nader, & Pruessner, 2013; Spring et al., 2015; Soeter & Kindt, 2010, 2011) because the drug typically takes around 90m to reach peak plasma concentration in the blood (Schiller & Phelps, 2011). The intention, therefore, is that the drug will be most potent during reconsolidation. However, in practice the drug will also be at least partially effective during the retrieval process itself. Additionally, in a series of transcranial magnetic stimulation (TMS) studies involving a finger-tapping sequence learning paradigm (see Chapter 3), the intervention was directly delivered *during* the reminder stage (Censor, Dimyan, & Cohen, 2010; Censor, Horovitz, & Cohen, 2014b; Censor, Dayan, & Cohen, 2014a). The subsequent performance impairment observed in the R^+I^+ group relative to the R^+I^- group could therefore be attributed to the latter group benefiting from an undisrupted reminder rather than the former group experiencing disruption of a reconsolidation process.

Meeting this criterion ($C2^Y$) simply requires that the reminder precedes the intervention. If it does not, the criteria has not been met ($C2^N$)

CRITERION 2: Reactivation precedes intervention

Reactivation must occur before rather than during or after the intervention.

2.1.4.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C2), and summary data are presented in Figure 2.2. 167 out of 168 cases met the criterion (99%). The remaining case did not meet it (1%).

Meeting this criterion is clearly much more straightforward when behavioural interventions are employed relative to slow-acting pharmacological interventions (see above). Indeed, our review revealed only a single case that did not meet the criterion, and this was an explicit attempt to identify if the observed effect was contingent on the temporal order of the reminder and intervention stages (Finn et al., 2012). Therefore, we can be confident that this literature almost entirely meets criterion 2.

2.1.5 Criterion 3: Baseline comparison

2.1.5.1 Description of criterion

Any reconsolidation-mediated modification of the target memory trace should be apparent in differential trace-dependent performance measured at baseline relative to final test (for the R^+I^+ condition). Baseline measures could be taken towards the end of training, or during the reminder stage if the reminder protocol involves active performance of the specific target material (see subsubsection 2.1.13.1). For example, when retrieval of a previously learned 'old' sequence was followed by learning a new sequence, Walker et al. (2003) observed a substantial drop in old sequence accuracy between the reminder stage and the final test stage, consistent with the prediction that the trace representing the old sequence had been compromised (cf. Hardwicke & Shanks, 2016, Chapter 3).

If a performance decline between baseline and final test is observed, then additional control conditions are required to ascertain whether the effect is dependent on the reminder and the intervention (C4, C5, C6, C7). Accounting for baseline performance allows one to ascertain whether any performance deficits in the R^+I^+ condition reflect absolute impairment (performance lower at final test compared to baseline) rather than just relative impairment (performance at final test in the R^+I^+ condition lower than in control conditions). For example, in a motor sequence-learning paradigm (as in Walker et al., 2003), Censor, Dayan, and Cohen (2014a) observed lower performance in a R^+I^+ group relative to a R^+I^- group on the final test. However, this was only a relative impairment: performance had actually improved from baseline to final test in both conditions, just to a lesser extent in the R^+I^+ group. Because there was no absolute impairment effect, these findings are not consistent with the prediction that post-retrieval new learning will disrupt reconsolidation of the reactivated memory trace causing its weakening or erasure.

It should be noted that absolute impairment effects can also be identified by including a condition that neither recieves reminder nor intervention (i.e., R^-I^-; C6). In fact, this approach has the advantage of controlling for any interim forgetting (unrelated to the key manipulations) that may have occurred between baseline and test. Nevertheless, measuring baseline performance is desirable to rule out the possibility that differential performance on the final test is due to initial inter-condition differences at baseline. Researchers often attempt to minimize baseline differences through the use of randomisation and/or counter-balancing procedures. However, due to unsystematic sources of error, some baseline differences are inevitable. A common strategy is to employ a statistical test to examine whether any baseline differences are reliable, and a separate statistical test to examine whether any final test differences are reliable. However, this implicitly assumes that the difference between significant and not significant is itself statistically significant, which is not necessarily the case (Gelman & Stern, 2006; Nieuwenhuis, Forstmann, & Wagenmakers, 2011). A small numerical difference in performance at baseline, even if not statistically significant, could contribute to a statistically significant difference on the final test. A more appropriate procedure therefore, is to directly compare baseline and final test effects, for example using a factorial ANOVA testing the time (baseline, test) by condition (e.g., R^+I^+, R^-I^-) interaction.

Cases are deemed to have met the present criterion ($C3^Y$) if they include a statistical comparison between baseline and final test performance *and* a significant difference in the expected direction[13] is observed (i.e., R^+I^+ baseline $\neq$ R^+I^+ test). If the comparison is conducted but there is no significant effect in the expected direction, then the case has not met the criterion ($C3^N$). If the comparison has not been conduced then the criterion cannot be assessed ($C3^?$).

[13]The direction of the effect will depend the nature of the intervention, and one's interpretation of reconsolidation theory (see subsubsection 2.1.13.3).

CRITERION 3: Baseline comparison

A statistical comparison between baseline and final test indicates a significant effect in the expected direction.

2.1.5.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C3), and summary data are presented in Figure 2.2. Only 7 out of 168 cases met the criterion (4%) while 12 cases did not (7%). Finally, 149 cases did not assess the criterion (89%).

The majority of cases did not include a statistical comparison between baseline and final test performance. The extent to which this is problematic will depend partly on whether the research design also includes a R^-I^- condition (i.e., C6), which can serve a similar function. However, adherence to C6 is also rare (see subsection 2.1.8). When C3 and C6 are not met, it is unclear whether any performance deficits observed in the R^+I^+ condition compared to controls reflected only relative impairment or also involved absolute impairment. It will be important to assess both criteria in future studies because reconsolidation-mediated disruption of the memory trace should be reflected in both relative and absolute impairment. Furthermore, even if C6 is met, it is important to also meet C3 in order to rule out the possibility that differences in baseline performance are unduly influencing final test outcomes.

2.1.6 Criterion 4: Reactivation-dependent

2.1.6.1 Description of criterion

It is well established that learning new information can have a detrimental impact on the subsequent recall of previously learned information (see ‘retroactive interference’ effects in Chapter 1). To attribute any modification effect to reconsolidation, it is necessary to demonstrate that the effect is dependent on the reactivation manipulation (Agren, 2014;

Nader & Hardt, 2009; Tronson & Taylor, 2007). If the effect occurs in the absence of reactivation then this suggests that it cannot be attributed to reconsolidation processes.

In practice, demonstrating the critical role of the reactivation manipulation involves comparing the R^+I^+ condition to a R^-I^+ condition (see Table 2.1). Often a between-subjects manipulation is used (e.g., J. C. K. Chan & LaPaglia, 2013; Hupbach et al., 2007; Wichert et al., 2011, also see Chapter 4 and Chapter 3), but some studies have employed a within-subjects manipulation by providing reminders for some stimuli, and not others (e.g., Pashler, Kang, & Mozer, 2013; Schiller et al., 2010, Chapter 5).

Cases are deemed to have met this criterion ($C4^Y$) if they include a R^-I^+ control group *and* a significant difference in the expected direction[13] is observed (i.e., $R^+I^+ \neq R^-I^+$). If the comparison is conducted, but there is no significant effect in the expected direction, then the case has not met the criterion ($C4^N$). If the required control group has not been included then the criterion cannot be assessed ($C4^?$).

CRITERION 4: Reactivation-dependent

Any modification effect must be dependent on the reactivation manipulation.

2.1.6.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C4), and summary data are presented in Figure 2.2. 36 out of 168 cases met the criterion (21%) while 65 cases did not (39%). Finally, 59 cases did not assess the criterion (35%) and for 8 cases (5%) it was unclear if it had been met or not (see notes in Table 2.5).

Whilst there are a concerning number of cases that did not assess this critical criterion, the majority of cases did include the critical R^-I^+ control group. For these cases, there are actually more instances in which a reconsolidation effect is *not* found. In other words, a key prediction of reconsolidation theory is not met. In most studies, cases in which there is no reconsolidation effect are typically accompanied by one or more cases in which it is claimed there is a reconsolidation effect (e.g., Beukelaar, Woolley, & Wenderoth, 2014; Soeter & Kindt, 2011; Wichert et al., 2011). Fewer studies report a complete absence of reconsolidation effects (e.g., Hardwicke, Taqi, & Shanks, 2016; Golkar, Bellander, Olsson, & Öhman, 2012; Pashler et al., 2013; Potts & Shanks, 2012). Studies that did observe a reminder-contingent effect will also need to obtain a critical reminder*intervention interaction (C7) if the effect is to be consistent with reconsolidation theory.

2.1.7 Criterion 5: Intervention-dependent

2.1.7.1 Description of criterion

It is well established that reminders themselves can influence behaviour. For example, the 'testing' effect refers to the phenomenon whereby attempting to retrieve information (as in a 'specific-active' reminder, see subsubsection 2.1.13.1) can facilitate subsequent performance (Roediger & Butler, 2011). Conversely, in a conditioning paradigm, unreinforced presentations of the CS^+ can impair subsequent performance (i.e., "extinction", see subsection 1.2.3, Bouton & Moody, 2004). Any modification effect must therefore also be critically dependent on the intervention manipulation (Nader & Hardt, 2009; Tronson & Taylor, 2007).

In practice, demonstrating the critical role of the intervention manipulation involves comparing the R^+I^+ condition to a R^+I^- condition (see Table 2.1). Both between-subjects (e.g., Beukelaar et al., 2014; Walker et al., 2003; Wichert et al., 2011), and within-subjects manipulations are used (e.g., J. C. K. Chan & LaPaglia, 2013; Hardwicke et al., 2016, Chapter 4). In some cases, the no-intervention condition involves skipping the intervention stage entirely (e.g., Wichert et al., 2011), whereas other studies involve placebo interventions that are missing some critical 'active' component of the main intervention. For example, Censor et al. (2010) employed TMS applied to a control site compared to TMS applied to the motor cortex, and Hupbach (2015) compared a group who learned a new word list to a group who completed Sudoko puzzles.

Cases are deemed to have met this criterion ($C5^Y$) if they include a R^+I^- control group *and* a significant difference in the expected direction[13] is observed (i.e., $R^+I^+ \neq R^+I^-$). If the comparison is conducted, but there is no significant effect in the expected direction, then the case has not met the criterion ($C6^N$). If the required control group has not been included then the criterion cannot be assessed ($C6^?$).

CRITERION 5: Intervention-dependent

Any modification effect must be dependent on the intervention manipulation.

2.1.7.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C5), and summary data are presented in Figure 2.2. 54 out of 168 cases met the criterion (32%) while 26 cases did not (15%). Finally, 86 cases did not assess it (51%) and for 2 cases (1%) it was unclear if the criterion had been met or not (see notes in Table 2.5).

Just over half of the reviewed cases did not assess this criterion and a small but not insignificant number found that the intervention was ineffective. In both types of case, it could be that any observed modification effect is being driven by the independent effects of the reminder. A reasonable number of studies did demonstrate that the effect was intervention-contingent. These studies will also need to obtain a critical reminder*intervention interaction (C7) if the effect is to be consistent with reconsolidation theory.

2.1.8 Criterion 6: Control for independent effects

2.1.8.1 Description of criterion

As noted above (C4, C5), both reminders and interventions can have independent effects on performance that are unrelated to reconsolidation. Evidence for reconsolidation will require that there is a reminder*intervention *interaction* (C7). In order to evaluate that interaction, a R^-I^- control group is required so that the independent effects of the reminder (R^+I^-) and intervention (R^-I^+) can be estimated. This is discussed further for criterion 7 below.

Cases are deemed to have met this criterion ($C6^Y$) if they include a R^-I^- control group *and* a significant difference in the expected direction[13] is observed (i.e., $R^+I^+ \neq R^-I^-$). If the comparison is conducted, but there is no significant effect in the expected direction, then the case has not met the criterion ($C6^N$). If the required control group has not been included then the criterion cannot be assessed ($C6^?$).

CRITERION 6: Control for independent effects

It is necessary to control for any independent effects of the reminder and intervention.

2.1.8.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C6), and summary data are presented in Figure 2.2. 30 out of 168 cases met the criterion (18%) while 7 cases did not (4%). Finally, 129 cases did not assess it (77%) and for 2 cases (1%) it was unclear if the criterion had been met or not (see notes in Table 2.5).

In a small subset of cases, the required control group was included and the expected difference from the R^+I^+ group obtained. Only in a few cases was the control group included and no difference observed. In the vast majority of cases this control was not included at all. The main implication of this is that in those cases the critical reminder*intervention interaction cannot be assessed (see C7 below).

2.1.9 Criterion 7: Reminder*intervention interaction

2.1.9.1 Description of criterion

As noted above, in order for any modification effect to be attributed to reconsolidation, it must be contingent on both the reminder manipulation (C4) *and* the intervention manipulation (C5). Both manipulations however, could have independent effects on performance (C6). If those independent effects summated, it could lead to a pattern of results that appears to support reconsolidation theory, but is actually caused by unrelated effects. For example, consider a study in which a non-reinforced CS^+ is used as a reminder, and repeated presentations of the CS^+ (i.e., extinction trials) are used as the intervention (e.g., Schiller et al., 2010, see subsection 2.2.3). The reminder could, by itself, lead to a reduction in performance on a subsequent test. Similarly, the intervention alone could also impair subsequent performance. Together then, this R^+I^+ condition would likely perform worse than the other experimental conditions (R^-I^+, R^+I^-, R^-I^-), simply due to the main effect of the reminder and the main effect of the intervention. It is therefore necessary to demonstrate that the performance decrement goes 'above and beyond' what can be attributed to the independent effects of the reminder and intervention. This requires a significant reminder*intervention interaction.

Cases are deemed to have met this criterion ($C7^Y$) if they test the critical reminder*intervention interaction *and* find that it is statistically significant. If the comparison is made, and the interaction is not significant, then the criterion is not met ($C7^N$). If the comparison is not made, then the criterion cannot be assessed ($C7^?$), unless the data are available for further analysis (e.g., see subsection 2.2.2, James et al., 2015).

CRITERION 7: Reminder*intervention interaction

It is necessary to demonstrate that any modification effect is caused by a reminder*intervention interaction

---

2.1.9.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C7), and summary data are presented in Figure 2.2. 3 out of 168 cases met the criterion (2%) while 15 cases did not meet it (9%). Finally, 142 cases did not assess it (85%) and for 8 cases (5%) it was unclear if the criterion had been met or not (see notes in Table 2.5).

In the vast majority of cases, the critical reminder*intervention interaction was not tested. In most cases, this was because the necessary experimental conditions were not available. However, even when those conditions were included in the design, the interaction was not always tested. For instance, in Chan and LaPaglia (2013, for further details see Chapter 4), reminder (R^+, R^-) was manipulated between-subjects and intervention was manipulated within-subjects (I^+, I^-), providing a fully crossed factorial design. The analysis approach involved separate statistical comparisons between R^+I^+ and R^+I^- (a statistically significant effect, confirming C5), and R^-I^+ and R^-I^- (no significant effect). This pattern of results is intuitively appealing because it appears that the intervention was only effective when a reminder preceded it. As a result, it was concluded that disruption of reconsolidation had occurred in the R^+I^+ group. However, this analysis approach is problematic because there was no statistical comparison between the reminder conditions (nor is there a direct test of C4). In other words, it was incorrectly assumed that a qualitative difference between statistical significance in the two separate conditions must also entail a meaningful quantitative difference between them, and that is not necessarily the case (Gelman & Stern, 2006; Nieuwenhuis et al., 2011). Such a conclusion would require a statistical comparison of the difference scores, or perhaps a 2x2 factorial ANOVA to evaluate the reminder*intervention interaction.

Of course, just because the statistical comparison was not reported does not mean that it would not be significant, but this cannot be ascertained from the information provided in the article. In addition, even in the absence of a significant interaction, the outcome would be ambiguous, rather than evidence against the reconsolidation hypothesis per se (e.g., the power to detect a true interaction might be low). Nevertheless, the absence of a significant interaction makes for a rather unconvincing demonstration of reconsolidation. In subsection 2.2.2, a similar case (James et al., 2015) is discussed where a re-analysis of the publicly available dataset indicates that there is no significant interaction, undermining the claim that the observed effect was due to disruption of reconsolidation.

2.1.10 Criterion 8: Time-dependence of intervention

2.1.10.1 Description of criterion

Several non-human animal studies suggest that the effectiveness of a post-retrieval intervention is time-dependent (e.g., Judge & Quartermain, 1982; Mactutus et al., 1982; Nader et al., 2000a; Przybyslawski & Sara, 1997), mirroring findings with post-acquisition interventions in consolidation studies (see subsection 1.2.7). This is consistent with the idea that reconsolidation involves a progressive restabilisation process (Nader & Hardt, 2009; Tronson & Taylor, 2007). An ideal assessment of this prediction would involve systematic manipulation of the reminder-intervention interval. This strategy has been used effectively in consolidation studies (e.g., Duncan, 1949, see Figure 1.7) to expose the anticipated temporal gradient.

In practice, the time-dependent effectiveness of the intervention is typically demonstrated by comparing outcomes in just two conditions (Tronson & Taylor, 2007): an immediate (or short delay) post-retrieval intervention condition (i.e., standard R^+I^+), and a 6h delayed post-retrieval intervention (i.e., R^+I^D, see Table 2.1). In the prominent Nader et al. (2000a) study for instance, a PSI was found to induce profound amnesia when delivered immediately after the reminder, but was ineffective when delivered after a 6h delay (see subsection 1.2.8). This suggests that 6h was an upper bound on the reconsolidation window, at least given the parameters used in that particular study. It is unclear to what extent the 6h upper bound will generalise to other studies, particularly behavioural as opposed to pharmacological interventions. Nevertheless, a 6h reminder-intervention interval seems to have become a generally accepted ‘rule-of-thumb’ (Tronson & Taylor, 2007).

Cases are deemed to have met this criterion ($C8^Y$) if they examined performance in at least one delayed intervention condition (R^+I^D) *and* found that the effectiveness of the intervention declined with an increasing reminder-intervention interval. If an R^+I^D group was included, but the effectiveness of the intervention was not found to be time-dependent, then the criterion is not met ($C8^{N}$). If a $R^{+}I^{D}$ group was not included then the criterion cannot be assessed ($C8^{?}$).

CRITERION 8: Time-dependence of intervention

It is necessary to demonstrate that the effectiveness of the intervention is time-dependent.

2.1.10.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C8), and summary data are presented in Figure 2.2. 2 out of 168 cases met the criterion (1%) while 6 cases did not meet it (4%). 96 cases did not assess it (57%) and for 2 cases (1%) it was unclear if the criterion had been met or not (see notes in Table 2.5). For 62 cases (37%) the criterion was no longer applicable because a reconsolidation effect was not observed in the main experimental groups (C3, C4, C5, C6, C7).

The time-dependent effectiveness of the intervention is a key prediction of reconsolidation theory. However, very few studies have assessed this criterion. In early consolidation studies that employed behavioural interventions, it was the absence of a clear temporal gradient that led psychologists to shift their focus to response competition theories (see subsection 1.2.1 and subsubsection 1.2.7.4; J. A. McGeoch, 1942; Wixted, 2004). It is now important that the temporal gradient be evaluated in reconsolidation studies in order to ascertain whether a similar retrieval-based explanation is applicable here too.

2.1.11 Criterion 9: Time-dependence of test

2.1.11.1 Description of criterion

One particularly intriguing finding that emerged from the consolidation literature was that experimentally-induced amnesia tended not be apparent on short-term memory (STM) tests, but emerged later on long-term memory (LTM) tests (see subsubsection 1.2.7.6). In studies using protein-synthesis inhibitors (PSIs), it was argued that this pattern was expected by a dual-trace consolidation theory (McGaugh & Dawson, 1971). Because the PSI would have no influence on the electrophysiological activity associated with the STM trace, but would disrupt protein-synthesis necessary for the formation of the LTM trace, performance should remain unaffected after a short delay because the STM trace is intact. However, in studies using electroconvulsive shock (ECS), which *does* disrupt the STM trace, the same pattern was observed (e.g., Geller & Jarvik, 1968b; McGaugh & Landfield, 1970; R. R. Miller & Springer, 1971). It is not clear how consolidation theory, and by extension reconsolidation theory, can account for this finding (see subsubsection 1.2.7.6). Nevertheless, reconsolidation investigations in non-human animals have predominantly relied on PSIs as their choice of intervention (Besnard et al., 2012; Tronson & Taylor, 2007) and the finding of intact STM, but impaired LTM, is often taken as compelling evidence that a reconsolidation process has been disrupted (e.g., Nader et al., 2000a). The extent to which this pattern will generalise to studies using behavioural interventions is unclear.

Typical practice is to conduct an STM test shortly after termination of the intervention, and conduct an LTM test after a 24h interval (Tronson & Taylor, 2007). As with C1 (see subsection 2.1.3), the choice of a 24h interval is fairly arbitrary, but appears to meet most generally accepted upper-bounds on the completion of cellular consolidation (R. R. Miller & Springer, 1973).

Cases are deemed to have met this criterion ($C9^{Y}$) if they examined performance on both an STM-test ($R^{+}I^{+}T^{STM}$) and an LTM test (i.e., a standard $R^{+}I^{+}$ group) *and* found that performance was unaltered on the former, but modified on the latter. If the two conditions were compared, but showed no difference, then the criterion is not met ($C9^{N}$). If the two conditions are not included then the criterion cannot be assessed ($C9^{?}$).

CRITERION 9: Time-dependence of test

It is necessary to demonstrate that performance is unaffected shortly after the

intervention, but modified on a delayed (24h) test.

2.1.11.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C9), and summary data are presented in Figure 2.2. 5 out of 168 cases met the criterion (3%) and 7 cases did not (4%). 92 cases did not assess the criterion (55%) and for 0 cases (0%) it was unclear if it had been met or not (see notes in Table 2.5). For 64 cases (38%) the criterion was no longer applicable because a reconsolidation effect was not observed in the main experimental groups (C3, C4, C5, C6, C7).

Very few studies have examined this criterion. In studies that have examined and met it, the differential STM-LTM pattern is considered a hallmark of a reconsolidation effect (Hupbach et al., 2007; Schiller & Phelps, 2011; Walker et al., 2003). However, it is not logically clear whether unaffected STM should be expected when behavioural interventions are used because they could cause short-term interference effects that are unrelated to reconsolidation. Further theory development is needed to define exactly how behavioural interventions are thought to interact with the reconsolidation process.

2.1.12 Criterion 10: Absence of Recovery

2.1.12.1 Description of criterion

As outlined in Chapter 1, evidence from studies using recovery protocols (reinstatement, renewal, spontaneous recovery) indicates that behavioural interventions, such as extinction, do not result in the overwriting of existing memory traces (e.g., see subsection 1.2.3; Bouton, 2002). Furthermore, recovery from experimentally-induced amnesia can be observed even when more potent post-acquisition interventions are employed (e.g., PSIs, ECS, see subsubsection 1.2.7.8; R. R. Miller & Springer, 1973; Riccio & Richardson, 1984). Recovery effects appear to be inconsistent with the prediction that disruption of consolidation will lead to irreversible impairment of the target trace. However, some researchers have defended consolidation theory by arguing that recovery effects indicate that trace stabilisation was only partly disrupted, leaving a partial, recoverable trace (see paragraph 1.2.7.8.2; Cherkin, 1972; Gold & King, 1974; Squire, 2006). Nonetheless, observing recovery on a subsequent test makes any previous claim about trace modification seem less compelling. Arguably, recovery effects would be even more damaging for the theory that memory traces can be updated or overwritten via new information learned during the reconsolidation window, because this claim seems to imply complete destruction of the 'old' information (J. C. K. Chan & LaPaglia, 2013; J. L. Lee, 2009; Schiller et al., 2010).

Cases are deemed to have met this criterion ($C10^{Y}$) if they examine performance in a recovery protocol (e.g., reinstatement, renewal, or spontaneous recovery) *and* find that the modification effect observed during the first test stage has persisted. If a recovery protocol is used, and recovery is observed, then the criterion not met ($C1^{N}$). If a recovery protocol is not used then the criterion cannot be assessed ($C10^{?}$).

CRITERION 10: Absence of recovery

It is necessary to demonstrate that any modification effect persists, and there is no recovery of trace-dependent performance.

2.1.12.2 Evaluation of literature

Detailed case-by-case coding for this criterion can be found in Table 2.5 (column C10), and summary data are presented in Figure 2.2. 5 out of 168 cases met the criterion (3%) and 15 cases did not (9%). 97 cases did not assess it (58%) and for 1 case (1%) it was unclear if the criterion had been met or not (see notes in Table 2.5). For 50 cases (30%) the criterion was no longer applicable because a reconsolidation effect was not observed in the main experimental groups (C3, C4, C5, C6, C7).

Very few studies have examined whether observed modification effects can be recovered using one or more recovery protocols. For the studies that have employed a recovery protocol, the

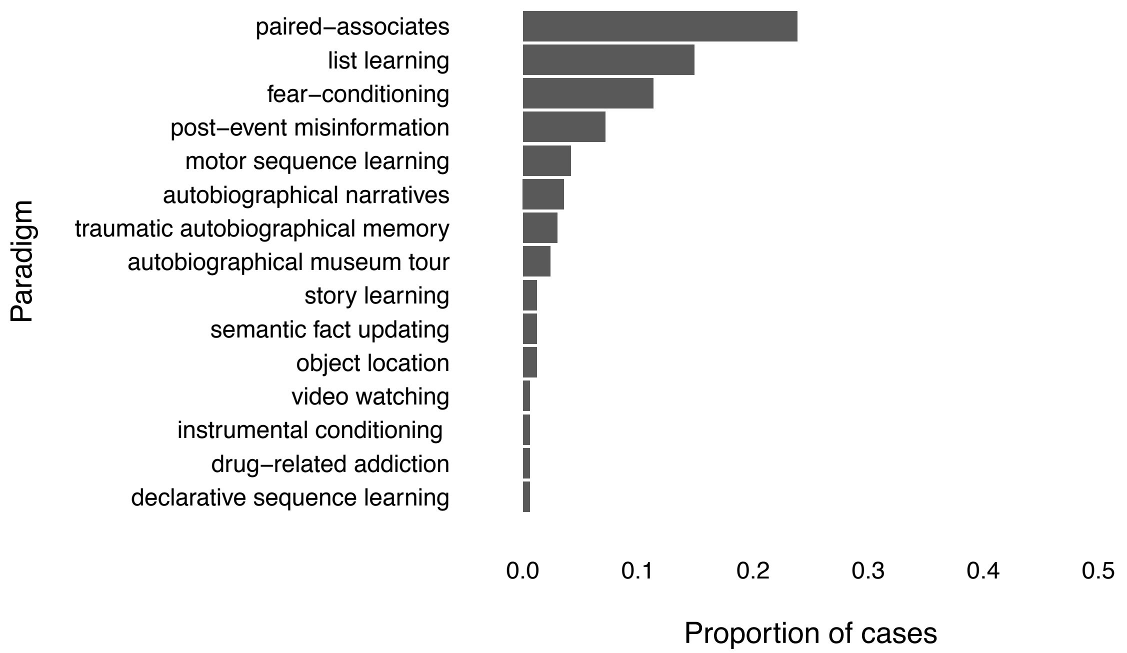


Figure 2.3: Proportion of cases using different types of paradigm in human reconsolidation studies. Cases refers to individual experimental conditions in which a reconsolidation effect is expected to occur, see Table 2.5.

majority (75%) observed recovery effects. Although there are relatively few cases to consider, the findings mirror a number of similar demonstrations of recovery following more potent interventions in non-human animal research on reconsolidation (e.g., Anokhin & Tiunova, 2002; Eisenberg & Dudai, 2004; Gisquet-Verrier et al., 2015; Lattal & Abel, 2004; Trent, Barnes, Hall, & Thomas, 2015). The implication of these findings is discussed in greater detail later in this chapter (see section 2.3).

2.1.13 General characteristics of the field

In this section we will outline some general characteristics of the field that emerge from the data gathered during the systematic review. The prevalence of different paradigms and intervention types is shown in Figure 2.3 and Figure 2.4 respectively. Further details are provided in the in-depth review of selected studies later in this chapter (section 2.2). Below we briefly discuss different types of reminder (subsubsection 2.1.13.1), distribution of sample sizes (subsubsection 2.1.13.2), and types of reconsolidation effect (subsubsection 2.1.13.3).

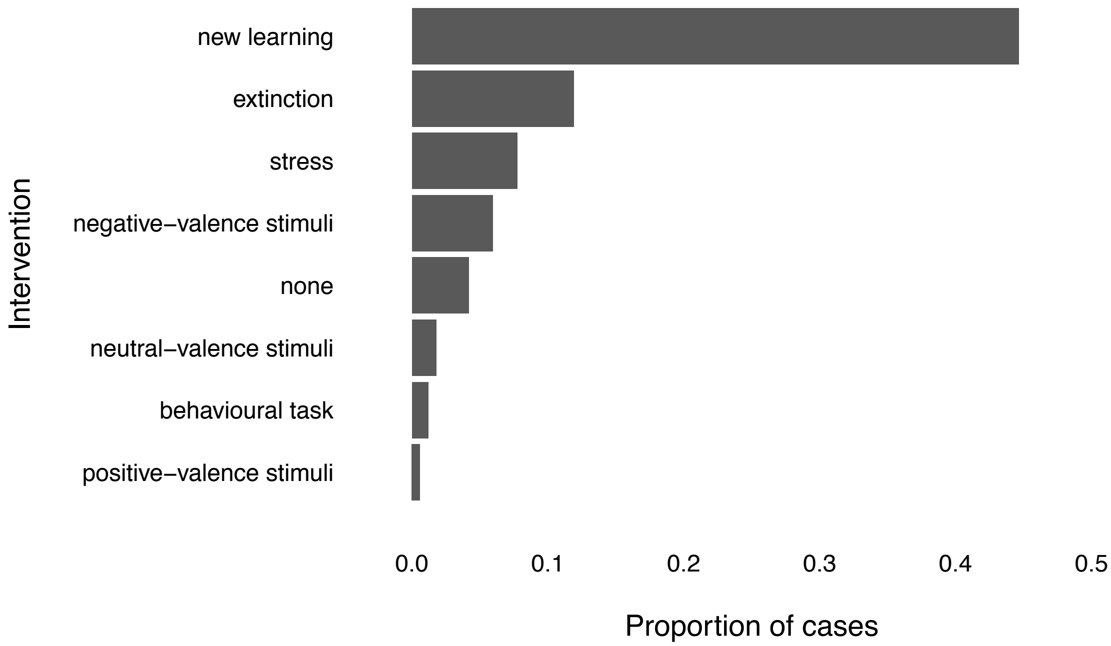


Figure 2.4: Proportion of cases using various types of behavioural interventions in human reconsolidation studies. Cases refers to individual experimental conditions in which a reconsolidation effect is expected to occur, see Table 2.5.

2.1.13.1 Reminder types

The nature of the Reminder Stage varies considerably between studies and we have broadly classified them into four types for the purposes of this review (see Figure 2.5).

Specific reminders refer to those in which the learning material from the Training Stage is either re-presented to the participant (Specific-*passive* reminder; S-P) or retrieved by them (Specific-*active* reminder; S-A). For example, during the Reminder Stage of Walker et al. (2003), participants re-performed a motor sequence task that they had learned the previous day (S-A reminder). By contrast, during the Reminder Stage of St. Jacques and Schacter (2013), participants were shown photographs taken during a museum tour they had experienced the previous day (S-P reminder).

General reminders do not refer to the specific learning material, but involve retrieval cues related to the broader contextual features of the original learning episode. In some cases, this can simply involve the participant receiving the Intervention Stage in the same physical context as the Training Stage (a General-*passive* reminder, e.g., Hupbach et al., 2008; B.

J. Jones et al., 2015). Similarly, in a study by Diekelmann et al. (2011), participants were re-exposed to an odour that had also been presented during the Training Stage, but was irrelevant to the learning task. In other cases, researchers have asked participants to think about general aspects of the learning task (e.g., 'Do you remember what we did yesterday?'), whilst instructing them not to recall the learning material itself (a General-*active* reminder, e.g., Hupbach et al., 2007).

The precise construction of the reminder is currently the subject of contention in the field (Forcato et al., 2009; Hupbach, 2015; Potts & Shanks, 2012). Many researchers have used S-A reminders because they have many practical advantages (e.g., J. C. K. Chan & LaPaglia, 2013; Hardwicke et al., 2016; Walker et al., 2003). Firstly, they indicate that the material learned on Day 1 was successfully encoded, and can now be successfully retrieved. Secondly, they provide an important manipulation check as they indicate which item representations were successfully reactivated, and thus presumably underwent reconsolidation (Chapter 4, J. C. K. Chan & LaPaglia, 2013). This means that a baseline level of trace-dependent performance can be established immediately prior to delivery of the intervention (see subsection 2.1.5).

However, some have proposed that the use of an S-A reminder is problematic because it involves *testing* participants' recall (J. C. K. Chan, 2009; Hupbach, 2015). As retrieval practice is known to aid subsequent recall (the "testing effect", Roediger & Butler, 2011), it could be that S-A reminders 'mask' other modification effects that would be attributable to reconsolidation (J. C. K. Chan, 2009; Hupbach, 2015). W. C. Gordon (1983) expressed a similar idea when discussing the design of early investigations into cue-dependent amnesia. He suggested that the cueing treatment should be designed in such a way as to allow for successful memory retrieval, but without providing an opportunity for additional learning. One counter-argument to this is that even following an S-A reminder, the trace presumably still needs to re-stabilise. Therefore, the deployment of an effective intervention should still disrupt the reconsolidation process, even if the S-A reminder led to additional learning. Resolution of this debate will require a more precise specification of the mechanisms that are thought to trigger destabilisation of the trace and the process by which new learning interacts with the reconsolidation process. The idea that 'prediction error' is a key feature of

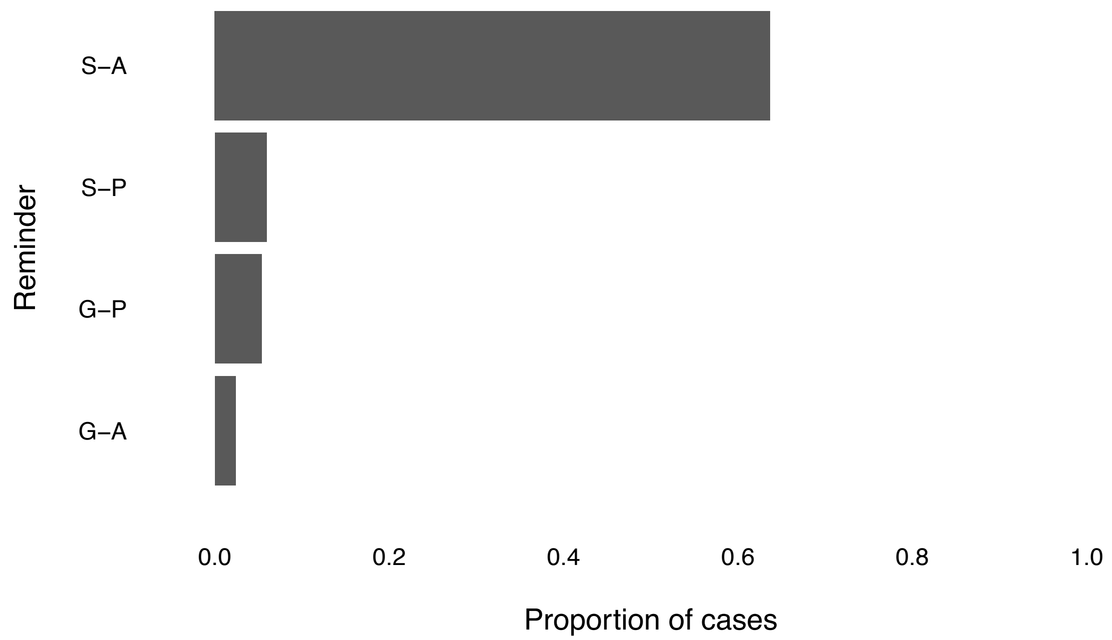


Figure 2.5: Proportion of cases using various types of reminder in human reconsolidation studies. *S-A* = Specific-active. *S-P* = Specific-passive. *G-P* = General-active. *G-A* = General-active. Cases refers to individual experimental conditions in which a reconsolidation effect is expected to occur, see Table 2.5.

a successful reminder intervention is discussed in section 2.3.

2.1.13.2 Sample sizes

The sample size of each critical R^+I^+ group was recorded and a histogram of these data is shown in Figure 2.6. Clearly there is a large skew towards the lower-end of the scale, with a large number of cases employing samples between 10-15 participants. Sample sizes ranged from 6 to 107 participants (median = 14). Small sample sizes are problematic because study outcomes are more susceptible to the influence of random variation. Consequently, small sample sizes produce less reliable estimates of the effect size (Lakens & Evers, 2014), have less chance of detecting true effects when they exist (J. Cohen, 1992), and increase the likelihood of false discoveries (Button et al., 2013; Ioannidis, 2005).

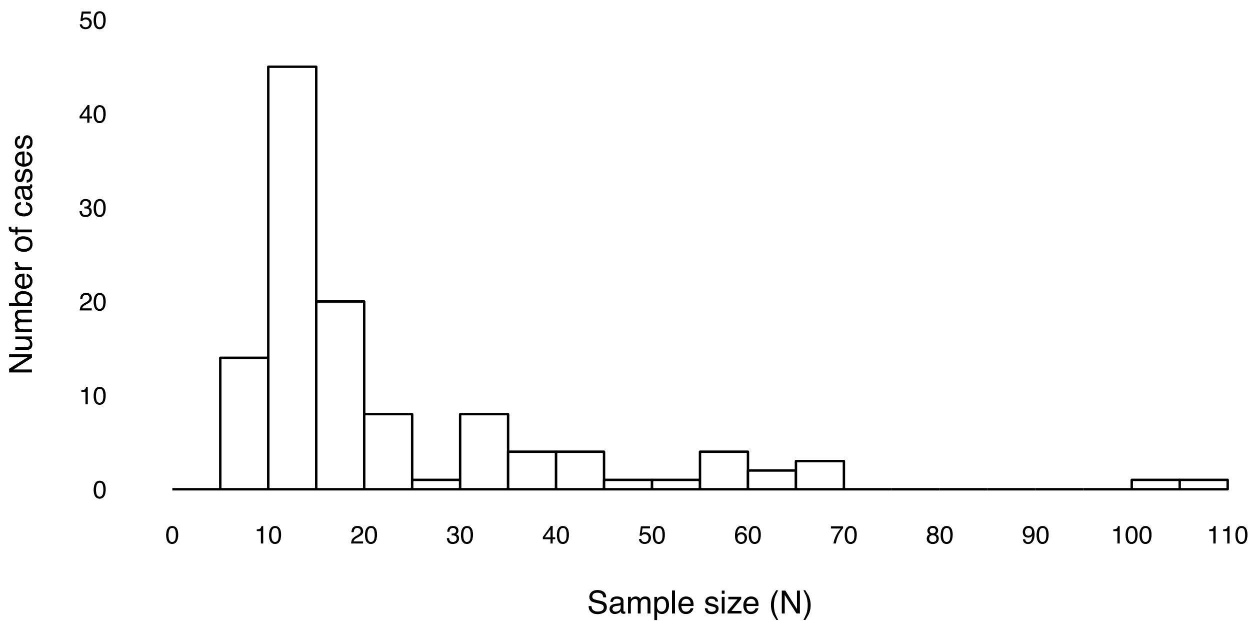


Figure 2.6: Histogram of sample sizes used in human reconsolidation studies employing behavioural interventions (bin width = 5).

2.1.13.3 Memory effects

Figure 2.7 shows the four types of effects that have been claimed in the reviewed cases. Destructive-updating effects (see Nader, 2003a) refer to those in which it is suggested that the intervention has disrupted the reconsolidation process and caused impairment to, or overwriting of, the restabilising trace (e.g., J. C. K. Chan & LaPaglia, 2013; Forcato et al., 2007; Schiller et al., 2010; Walker et al., 2003; Wichert et al., 2011). Conversely, constructive-updating effects refer to those in which new information learned during reconsolidation is ‘incorporated into’ the restabilising trace, without necessarily causing destructive loss of information (e.g., Hupbach et al., 2007; B. J. Jones et al., 2015; St Jacques & Schacter, 2013). Strengthening effects refer to the idea that an existing memory trace (see Sara, 2000b) can be ‘strengthened’ if reconsolidation is allowed to proceed unperturbed (e.g., Forcato et al., 2011), or enhanced via a stressful intervention (Finn & Roediger, 2011). These ideas are explored in more detail in section 2.2. Finally, in many cases, the researchers did not observe a reconsolidation effect.

For a literature that is generally interpreted as demonstrating strong support for reconsolida-

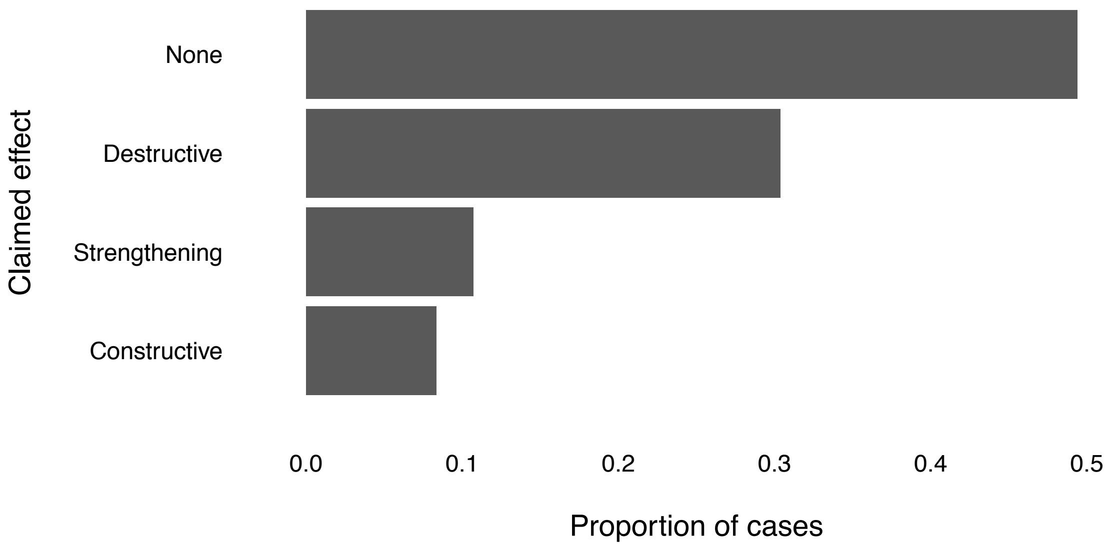


Figure 2.7: Proportion of cases claiming different types of effect in human reconsolidation studies. Cases refers to individual experimental conditions in which a reconsolidation effect is expected to occur, see Table 2.5.

tion theory (Agren, 2014; Besnard et al., 2012; Hardt et al., 2010; Schiller & Phelps, 2011; Schwabe et al., 2014), there are a surprisingly large number of cases in which a reconsolidation effect was not observed. Note that all cases refer to a R^+I^+ condition in which one might *a priori* anticipate a reconsolidation effect according to the core tenets of the theory. However, most *studies*, which can include several cases, tend to report at least one case that is found to elicit a reconsolidation effect. For example, one study reported evidence for reconsolidation when neutral-valence learning stimuli were used, but not when negative-valence or positive-valence stimuli were used (Schwabe & Wolf, 2009). Similarly, Walker et al. (2003) reported a reconsolidation effect in a motor-sequence learning task for the 'accuracy' dependent measure, but not for 'speed'. It is less common to fail to obtain reconsolidation effects outright, but there are several examples (e.g., Golkar et al., 2012; Hardwicke et al., 2016; Kindt & Soeter, 2013; Pashler et al., 2013; Potts & Shanks, 2012).

2.1.14 Interim summary

This systematic review of human reconsolidation investigations using behavioural interventions has revealed that few cases meet important criteria necessary for a robust and compelling

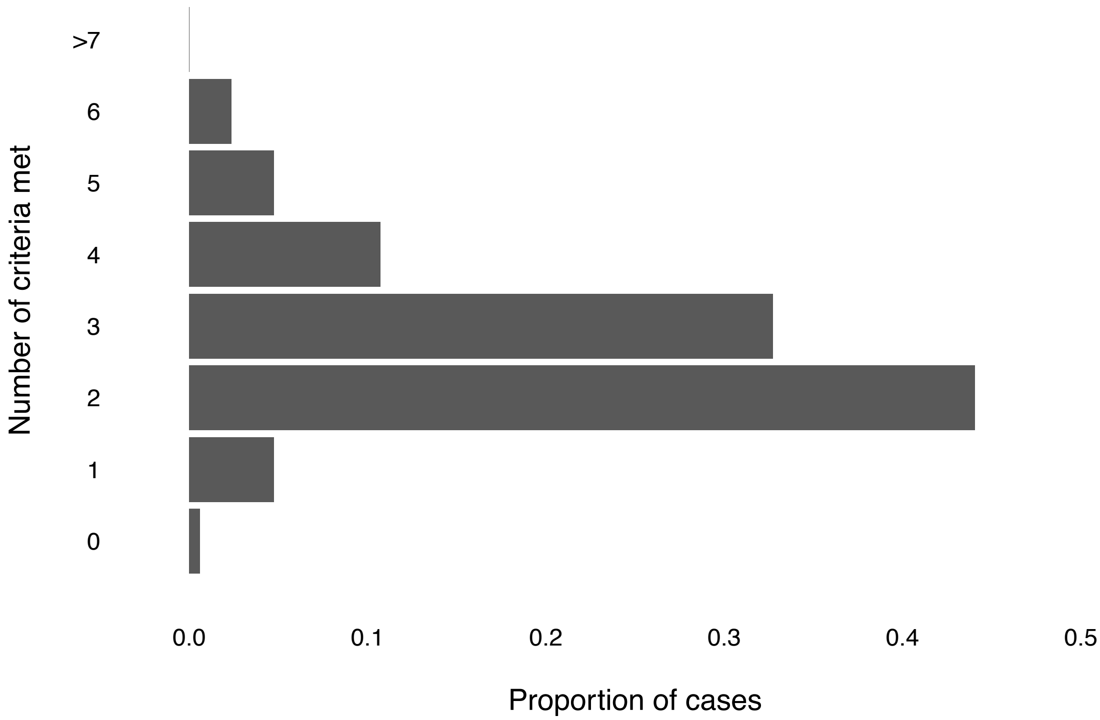


Figure 2.8: Proportion of cases claiming a reconsolidation effect that meet reconsolidation criteria (see Table 2.2). Cases refers to individual experimental conditions in which a reconsolidation effect is expected to occur, see Table 2.5.

demonstration of a reconsolidation effect (see Figure 2.2). In fact, out of 168 reviewed cases, most only meet three criteria or fewer (82%), and no case meets more than six criteria (see Figure 2.8). Even if one considers broad research programmes in which multiple studies have been conducted using essentially the same core paradigm, the number of criteria met overall does not exceed 6 and is typically lower (Figure 2.9). Therefore, there has arguably not been a single compelling demonstrating of memory reconsolidation using a behavioural intervention in human participants. In section 2.2, we provide a more in-depth examination of several key studies that are often taken to provide compelling demonstrations of reconsolidation-mediated updating.

2.2 Selective case study review

This section will provide an in-depth, selective review of key case studies in the human reconsolidation literature in order to demonstrate application of the reconsolidation criteria

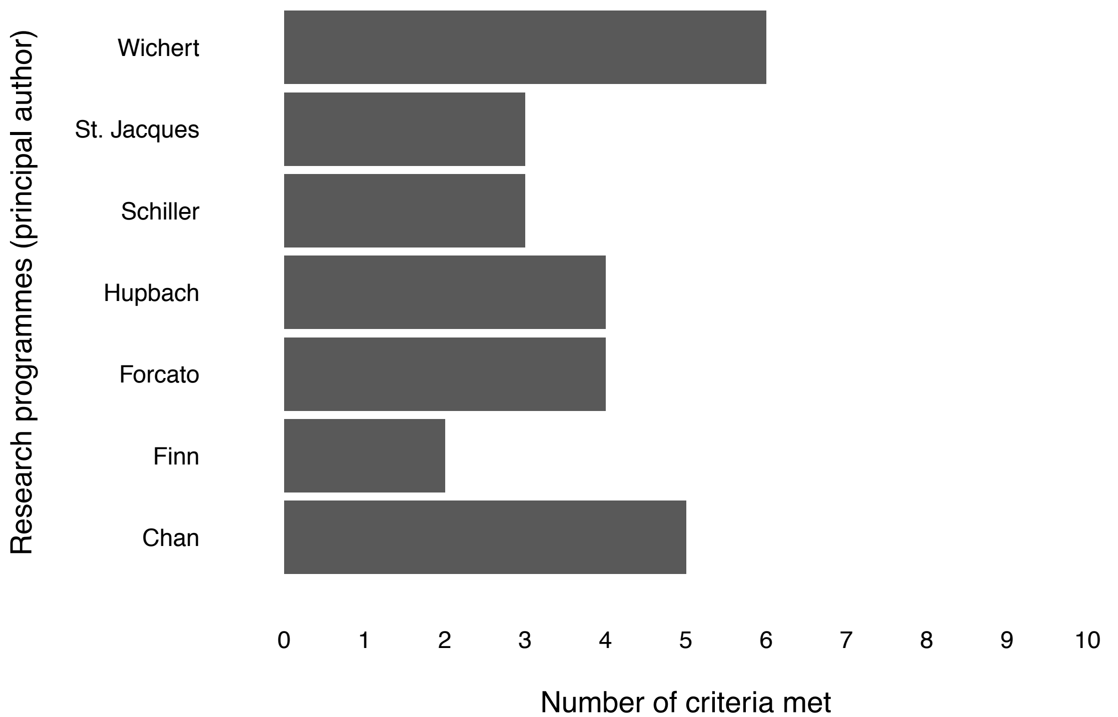


Figure 2.9: Number of reconsolidation criteria met by different research programmes.

in specific cases. These case studies have been selected because they are generally regarded as providing compelling demonstrations of reconsolidation in major review articles (e.g., Agren, 2014; Nader & Hardt, 2009; Schiller & Phelps, 2011; Schwabe et al., 2014) and illustrate how qualitative outcomes can vary quite substantially between studies (e.g., constructive updating and destructive updating). Firstly, we examine the claim that post-retrieval learning of a new list of objects can lead to reconsolidation-mediated constructive updating of the trace representing a previously learned list of objects (subsection 2.2.1, Hupbach et al., 2007). Secondly, we examine evidence that non-specific behavioural interference (playing *Tetris*) can disrupt the reconsolidation of traces representing experimentally-induced traumatic memory (subsection 2.2.2, James et al., 2015). Finally, we will turn to studies proposing that conducting extinction trials within the reconsolidation window can lead to destructive updating of ‘fear-memory’ in a conditioning paradigm (subsection 2.2.3, Agren et al., 2012; Björkstrand et al., 2015; Schiller et al., 2010). Discussion of two further cases (J. C. K. Chan & LaPaglia, 2013; Walker et al., 2003) is deferred to Chapters 3 and 4 respectively. Code for any re-analyses is available on the Open Science Framework (https://osf.io/m9und/).

2.2.1 Constructive-updating in list learning studies

In a series of studies, Hupbach and colleagues (Hupbach et al., 2007, 2008, 2009, 2011) investigated whether memory traces supporting declarative recall of object 'lists' could be updated in a reconsolidation paradigm. The general procedure (see Hupbach et al., 2007) was as follows: on Day 1, participants completed a Training Stage in which they were asked to memorise 20 assorted objects (List 1; e.g., balloon, calculator, sock etc.) presented sequentially by an experimenter and placed into a blue basket. After all of the objects had been shown and taken out of sight, the participant was asked to recall as many of them as possible. This was repeated until the participant could recall at least 17 out of 20 objects or 4 learning cycles had been completed[14]. 48h later on Day 2, participants completed a Reminder Stage in which they were taken to the same room as used on Day 1 by the same experimenter, shown the empty blue basket and asked "Do you remember this basket and what we did with it?" Participants were encouraged to describe the general learning procedure but not to mention specific objects. This was intended to be a 'subtle reminder' that reactivated the memory trace representing Day 1 information. Following the Reminder Stage participants completed an Intervention Stage in which they were asked to memorise 20 unrelated objects (List 2) placed simultaneously on a table in front of them for 30s (the learning procedure was intentionally different to Day 1 in order to reduce the likelihood that it would itself act as a reminder). The objects were then removed and the participant asked to recall as many of them as possible. As with Day 1 learning, this process was repeated until the participant could recall at least 17 out of 20 objects or 4 learning cycles had been completed. 48h later on Day 3, all participants completed a Test Stage in which they were asked to recall as many objects as possible *from List 1.* They were given four opportunities to do this and the

[14]Participants who did not reach the criterion were given a learning cycles score of 5 and the mean number of trials taken to reach this criterion is reported. However, the actual performance levels and number of participants who did not reach the criterion are not reported. This is important because differences in baseline performance are likely to influence performance on the final test, potentially obscuring the true impact of the independent variables (see subsection 2.1.5).

researchers reported the mean number of List 1 objects recalled[15] and the mean number of List 2 objects recalled (i.e., intrusions).

In Experiment$_{\text{HN07}}$ 1, there were three groups: a Reminder and Intervention Group who followed the above procedure (R^+I^+, $n = 12$), a No-Reminder Group who did not complete the Reminder Stage and completed the Intervention Stage in a different room with a different experimenter (R^-I^+, $n = 12$), and a No-Reminder, No-Intervention Group who skipped the Reminder Stage and Intervention Stage entirely (R^-I^-, $n = 12$). In Experiment$_{\text{HN07}}$ 2, a Reminder and Intervention STM-Test Group ($R^+I^+T^{STM}$, $n = 12$), No-Reminder STM-Test Group ($R^-I^+T^{STM}$, $n = 12$), and No-Reminder, No-Intervention STM-Test Group ($R^-I^-T^{STM}$, $n = 12$) completed the Test Stage immediately after the Intervention Stage on Day 2. Finally, in Experiment$_{\text{HN07}}$ 3, an additional Reminder and Intervention Group (R^+I^+T-L2, $n = 12$) and No-Reminder Group (R^-I^+T-L2, $n = 12$) were asked to recall List 2 rather than List 1 objects.

The findings were as follows. In Experiment$_{\text{HN07}}$ 1, the number of List 1 objects recalled during the Test Stage did not differ significantly across the R^+I^+, R^-I^+, and R^-I^- groups suggesting that the post-retrieval new learning of List 2 had not disrupted reconsolidation of the memory trace representing List 1[16]. However, the number of intrusions from List 2 was significantly larger in the R^+I^+ group ($C2^Y$; $C3^+$) compared to the R^-I^+ ($C4^Y$) and R^-I^- groups ($C5^Y$; there was no R^+I^- group: $C6^?$). This led the researchers to conclude that "the reminder did reactivate memory of the original list, returning it to a state in which new information could be incorporated" (Hupbach et al., 2007, p. 49).

In Experiment$_{\text{HN07}}$ 2, the number of List 1 objects recalled on the STM-Test was substantially

[15]The rationale for reporting the mean across four trials is unclear and it may underestimate participants' knowledge by diluting higher performance trials with lower performance trials. Indeed, a significant linear trend is reported indicating that recall improved across trials. As amnesia is a key prediction of reconsolidation theory it would arguably be more useful to report the maximum number of List 1 objects recalled because this would provide a more sensitive estimate of participants' knowledge.

[16]It should be noted that, numerically, List 1 recall was substantially (~8.7-13.2%) lower in the R^+I^+ group relative to the R^-I^+ and R^-I^- groups. It is possible that the experiment was not sufficiently powered to detect a true impairment effect.

lower in the groups receiving the intervention ($R^+I^+T^{STM}$, $R^-I^+T^{STM}$) compared to the group that did not ($R^-I^-T^{STM}$). However, the effect was not dependent on the reminder, as there was no significant difference between the $R^+I^+T^{STM}$ and $R^-I^+T^{STM}$ groups (i.e., there appears to have been a retroactive interference effect, but it was not attributable to reconsolidation). In contrast to the Day 3 test in Experiment_{HN07} 1, the number of intrusions from List 2 was practically non-existent across all STM-Test groups ($C9^Y$). Finally, in Experiment_{HN07} 3 when participants were asked to recall objects from List 2, there was no difference between the R^+I^+T-L2 and R^-I^+T-L2 groups in the number of List 2 objects or List 1 objects (intrusions) recalled. The time-dependence of the intervention ($C8^?$) and absence of recovery ($C10^?$) criteria were not evaluated in this study.

In Hupbach et al. (2008), the researchers ran a series of additional R^+I^+ groups (as in Experiment_{HN07} 1; $n = 12$ per group) in which they manipulated aspects of the Reminder Stage. Elevated intrusions were only observed when participants completed Day 2 in the same context as training, and was absent when participants completed Day 2 in a novel context, but met with the same experimenter and/or received the 'subtle reminder' question ($\text{Experiments}_{HN08}$ 1 and 2). Intriguingly, the effect was also absent when participants briefly entered the training context for the Reminder Stage, but were taken to a novel context for the Intervention Stage (Experiment_{HN08} 3). The researchers proposed that being returned to the encoding context for the Intervention Stage was both necessary and sufficient for the reconsolidation-mediated updating to occur (for further discussion of such 'boundary conditions' see section 2.3). Note that there were no No-Reminder[17] or No-Intervention

[17]B. J. Jones et al. (2015) conducted a replication of this experiment in young adults (M age = 19.8 years, SD = 2.8;) and older adults (age > 65 years; M and SD not provided) and employed Reminder (R^+I^+) and No-Reminder (R^-I^+) Groups. As in Hupbach et al. (2007) and Hupbach et al. (2008) recall of Set 1 objects did not vary substantially between R^+I^+ and R^-I^+ (for both age groups) and higher intrusion levels were observed in R^+I^+ relative to R^-I^+ (for young adults). However, unexpectedly the intrusion pattern was reversed in the older adults; intrusion levels in R^+I^+ were comparable to the young adult group, but lower than in R^-I^+. B. J. Jones et al. (2015) also observed higher intrusions rates in the R^+I^+ group relative to the R^-I^+ group in a sample of young rats performing an analogous task based on the Hupbach paradigm (also see B. J. Jones, Bukoski, Nadel, & Fellous, 2012). But the reminder had no substantial influence on intrusion levels in a group of older rats (B. J. Jones et al., 2015). The reasons for this pattern of results are

groups ($C4^?$, $C5^?$, $C6^?$) in this study and it is necessary to compare to the relevant control groups in Hupbach et al. (2007). Additional criteria were also not implemented/assessed ($C8^?$, $C9^?$, $C10^?$).

Hupbach et al. (2009) followed a similar procedure to Hupbach et al. (2007) but used an old/new recognition test and source memory test. During the Test Stage, the experimenter read out the names of List 1, List 2, and novel 'foil' objects in a randomised order (60 objects in total), and the participant had to indicate whether each object was 'old' or 'new'. If they responded 'old', the experimenter asked whether the object had been presented during the Training Stage ('Monday') or the Intervention Stage ('Wednesday'). There was a Reminder and Intervention Group (R^+I^+, $n = 10$) and a No-Reminder Group (R^-I^+, $n = 10$). Participants in both groups had comparable, almost perfect recognition performance for both List 1 and List 2 objects (>91% 'hits'), providing compelling evidence against any form of *destructive* updating effect. Participants in both the Reminder and No Reminder Group were also equally good at attributing List 1 objects to their correct source (>87% correct). However, participants in the Reminder group were more likely to attribute List 2 objects to List 1 relative to the No Reminder Group ($C4^Y$), echoing the asymmetrical intrusion effect previously observed in a free-recall task (Hupbach et al., 2007). Note that there was no No-Intervention group ($C5^?$, $C6^?$) in this study and it is necessary to compare to the relevant control groups in Hupbach et al. (2007). Additional criteria were also not implemented/assessed ($C8^?$, $C9^?$, $C10^?$).

Hupbach et al. (2011) also followed a similar procedure to Hupbach et al. (2007) and investigated the role of context familiarity in a sample of young children (~5 years old). The familiarity of a context was defined as whether participants had encountered it (children's home) or not (previously unvisited room in daycare center) prior to the experiment. Firstly (Experiment_{HN11} 1), the researchers replicated the previous finding (Hupbach et al., 2008) that completing Day 2 stages in the same context as training (R^+I^+, $n = 11$) was sufficient to elevate intrusion rates relative to a No-Reminder Group who completed Day 2 in a different spatial context (R^-I^+, $n = 9$; $C4^Y$). In a second experiment (Experiment_{HN11} 2),

unclear (B. J. Jones et al., 2015) and will require further empirical investigation.

intrusion effects were relatively low when the Day 1 and Day 2 context was the same, but pre-experimentally familiar (R^+I^+, $n = 12$). By contrast, elevated intrusions were observed in a group who received a 'three-component-reminder' (R^+I^+, $n = 12$) consisting of a familiar spatial context, the same experimenter as Day 1, and a 'subtle reminder' about the learning procedure from Day 1 (see above, Hupbach et al., 2007). Therefore, it would seem that the familiarity of the spatial context constrains whether the intrusion effect will occur. In addition, other 'reminders' (experimenter, subtle reminder), appear to become effective in a familiar context (cf. Hupbach et al., 2008).

Finally in $\text{Experiment}_{\text{HN11}}$ 3, all participants received a three-component reminder (as in $\text{Experiment}_{\text{HN11}}$ 2) but one group of children were asked to recall List 1 objects immediately after the Intervention Stage ($R^+I^+T^{STM}$, $n = 8$) and a second group were asked to recall List 2 objects on Day 3 (R^+I^+T-L2, $n = 8$). In comparison to the three-component reminder group of $\text{Experiment}_{\text{HN11}}$ 2, both of these groups had substantially lower intrusion rates, suggesting the effect was dependent on the time of testing ($C9^Y$) and asymmetric (as in Hupbach et al., 2007). Throughout this study (Hupbach et al., 2011) the 'updating effect' was restricted to differences in intrusion levels: there was no impairment of List 1 recall. In addition, there were no No-Intervention controls ($C5^?$, $C6^?$; it is necessary to compare to the relevant control groups in Hupbach et al., 2007) and other criteria were not implemented/assessed ($C8^?$, $C10^?$).

To summarise, in contrast to much of the non-human animal literature (e.g., Nader et al., 2000a), and a prior study which employed a post-retrieval new learning intervention in human participants (Walker et al., 2003), Hupbach and colleagues (Hupbach et al., 2007, 2008, 2009, 2011) did not observe a destructive updating effect (cue-dependent amnesia): recall of List 1 objects did not differ significantly across experimental groups[18]. This effect is clearly inconsistent with the predictions of reconsolidation theory as the new learning intervention does not appear to have perturbed the restabilisation of the List 1 memory trace. Nevertheless, the researchers did observe an intriguing asymmetric source-misattribution effect: receiving

[18]Although as mentioned above, it could be that these studies were not sufficiently powered to detect a true impairment effect.

a List 1 reminder 48h after the Training Stage ($C2^{Y}$), and immediately prior to learning List 2 ($C3^{+}$), led to an elevated number of intrusions/misattributions when participants were subsequently asked to recall (Hupbach et al., 2007, 2011) List 1 objects, or identify the source of List 2 Objects (Hupbach et al., 2009; also see Gershman et al., 2013), but hardly any intrusions/misattributions occurred when participants were asked to recall or identify the source of List 2 objects ($R^{+}I^{+}$ and $R^{+}I^{+}$T-L2 respectively). The effect was contingent on the reminder ($R^{-}I^{+}$; $C4^{Y}$), and the intervention ($R^{-}I^{-}$; $C5^{Y}$; although cross-experiment comparisons are necessary). The absence of a $R^{+}I^{-}$ group prevented a test of the critical reminder-intervention interaction ($C6^{?}$), but it seems highly unlikely that the intrusion effect could be driven solely by the reminder because production of List 2 items presumably relies on having learned List 2. Consistent with the delayed onset of reconsolidation effects observed in previous studies (e.g., Nader et al., 2000a; Walker et al., 2003), the intrusion effect was absent on a test conducted shortly after the Intervention Stage but emerged 48h later ($R^{+}I^{+}T^{STM}$; $C9^{Y}$). The time-dependence of the intervention ($C8^{?}$) and absence of recovery ($C10^{?}$) criteria were not evaluated in this series of studies.

Hupbach and colleagues have taken these findings to indicate that reconsolidation allows for the “incorporation of new information into old memories” and suggest that “memory ‘traces’ are dynamic rather than static” (Hupbach et al., 2007, p. 52). More broadly, this is a proposal that reconsolidation enables an updating process that is *constructive* (also see St Jacques & Schacter, 2013) rather than *destructive* (J. C. K. Chan & LaPaglia, 2013; Walker et al., 2003; Schiller et al., 2010, see Figure 2.7). Although intuitively appealing, the idea that new information can be ‘incorporated into a memory’ is poorly specified. The explanation is trying to capture the phenomenon that an instruction to recall items that belong to List 1 not only elicits responding with List 1 items, but also responding with List 2 items. Colloquially, participants’ ‘memory’ for List 1 has been altered in the sense that they now (appear to) think that List 2 items occurred on Day 1 when they learned List 1. However, in terms of the stage analysis presented in Chapter 1 (see section 1.1), it is participants’ *ecphoric memory* that contains a mixture of List 1 and List 2 items. It is by no means inevitable that this ecphoric phenomenon relies on physical modification of *memory traces* (see Figure 1.1). The observed effect could simply involve the formation of associative links between the original trace and a trace representing the new information. Alternatively, it could involve the formation of an entirely new memory trace representing old and new information that co-exists with the original trace. The theoretical construct of reconsolidation seems unnecessary in both scenarios.

More concretely, Sederberg, Gershman, Polyn, and Norman (2011; also see Gershman et al., 2013) argued that the critical asymmetric intrusion effect reported by Hupbach and colleagues (Hupbach et al., 2007, 2009) could be explained without recourse to a reconsolidation process via the established principles of contextual reinstatement and item–context binding instantiated in the Temporal Context Model (TCM; see Figure 1.1, orange triangle). In TCM, the temporal context is represented by a recency-weighted average of recently encountered items. When an item is encoded, it is ‘bound’ to the current temporal context. Items can be retrieved by cueing with a temporal context (e.g., ‘Recall Set 1 from Monday’). Successful item retrieval also retrieves the temporal context associated with that item from encoding, which in turn can facilitate the retrieval of other items associated with the same temporal context.

The TCM can account for the Hupbach et al. (2007) finding as follows: during the Training Stage (Day 1), in addition to learning list 1 objects, participants learn that those objects are associated with various aspects of the external environment (e.g., mode of presentation, experimenter, room) and the current temporal context (from herein ‘Context A’). In the Reminder Stage (Day 2), memory traces representing Context A elements will be reactivated. During the subsequent Intervention Stage, List 2 objects will become associated with the various aspects of the external environment and current temporal context (Context B) *in addition* to the reactivated Context A elements. Therefore, there is a critical asymmetry in the Reminder Group: List 2 is associated with Context A and Context B, whereas List 1 is only associated with Context A. Consequently, during the Day 3 Test Stage, cueing List 1 triggers recall of both List 1 objects and List 2 objects (i.e., intrusions) whereas cueing List 2 triggers only recall of List 2 objects. For participants in the No-Reminder Group, List 1 is only associated with Context A and List 2 is only associated with Context B, so intrusions are substantially lower.

Recall that Hupbach et al. (2007) also observed unaltered performance in a $R^{+}I^{+}T^{STM}$ group indicating that the observed intrusion effect was contingent on the time of test ($C9^{+}$). Sederberg et al. (2011) suggest that this can be attributed to participants use of a 'recall-to-reject' strategy to discriminate between List 1 and List 2 items. In response to the List 1 cue, the List 2 objects are successfully retrieved, but withheld (see Figure 1.1, conversion) because the retrieved context for those items (Context B) matches the current context (Context B), therefore indicating the object was seen 'today' (Day 2) and could not be a member of List 1 1 (Context A). Such a strategy would of course be invalid on the Day 3 Test because none of the retrieved objects would match the current context (Context C, i.e., Day 3). In summary therefore, although several key criteria of the reconsolidation theory appear to have been met in this series of studies, alternative perspectives that favour retrieval-based mechanisms are also plausible. Furthermore, it is important to recognise that there is no evidence of a destructive updating effect in any of these studies (List 1 recall is always intact), which is a qualitatively different outcome to that of non-human studies of reconsolidation using invasive interventions (Nader et al., 2000a), and many human studies that have used behavioural interventions (e.g., J. C. K. Chan & LaPaglia, 2013; Walker et al., 2003; Schiller et al., 2010).

2.2.2 General behavioural interference (Tetris)

Whilst some studies have employed new learning as a form of intervention (Walker et al., 2003; Hupbach et al., 2007), others have employed more general behavioural tasks, such as playing the computer game *Tetris* (James et al., 2015). In contrast to the idea that new learning might overwrite related content stored in existing traces (J. C. K. Chan & LaPaglia, 2013; Schiller et al., 2010), the approach adopted here appears to be more in line with Wixted's (2004) theory about behavioural tasks having non-specific disruptive effects on consolidation processes (see subsubsection 1.2.7.4). Specifically, James et al. (2015) were interested in the potential of post-retrieval Tetris to disrupt the reconsolidation of memory traces representing a traumatic event. Participants first viewed a short video depicting unpleasant scenes (e.g., car-pedestrian collision; Training Stage). 24h later on Day 2 ($C1^{Y}$), participants completed a Reminder Stage ($C2^{Y}$) which involved viewing a series of stills from the video. After a short

10m delay, they played *Tetris* for 12m (Intervention Stage). For the Test Stage, participants were asked to keep a daily diary until Day 7 in which they recorded any incidence of 'intrusive memory' (spontaneous recall) involving content from the video. Additionally, on the final day (Day 7) participants also completed visual and verbal true/false recognition tests for the video content.

In Experiment$_{JH}$ 1, the incidence of intrusive memories across the Day 1 - Day 7 period was significantly lower in a R^+I^+ group (n = 26) relative to a R^-I^- group (n = 26, $C6^Y$), but the manipulation had no impact on recognition test performance ($C6^N$). The researchers suggest that these findings provide tentative evidence for reconsolidation theory, proposing that playing Tetris leads to competition for 'working memory resources', resulting in 'cognitive blockade' of the reconsolidation process. Nevertheless, they acknowledge that additional R^+I^- and R^-I^+ groups are required to ensure the observed effect is not simply a result of the reminder and/or intervention acting independently ($C4^?$, $C5^?$, $C7^?$). The absence of an effect on the recognition test was given little attention even though it appears to contradict the key predictions of reconsolidation theory (i.e., impairment or erasure of underlying memory traces).

Experiment$_{JH}$ 2 involved the same procedure as above, but with the inclusion of aforementioned additional control groups (n = 18 per group), thus enabling a fully crossed factorial design. As in Experiment$_{JH}$ 1, there was no difference between groups on the recognition memory test. Additionally, the incidence of intrusive memories was significantly lower in the R+I+ group relative to the R-I- group ($C6^Y$; see Figure 2.10). It was also significantly lower than in the R+I- group ($C5^Y$) and R-I+ groups ($C4^Y$). However the statistical analysis employed did not allow for a test of the critical reminder*intervention interaction ($C7^?$). Specifically, the researchers employed a one-way ANOVA with Experimental Condition as a combined between-subjects factor, rather than a two-way factorial ANOVA with Reminder (R^+, R^-) and Intervention (I^+, I^-) as separate between-subject factors.

Fortunately, because the study data has been made openly available (https://osf.io/izxv2/) it is possible to run this alternative analysis and re-plot the individual participant scores (see Figure 2.10). The ANOVA indicates no main effect of Reminder, $F(1,68) = 2.315$, p

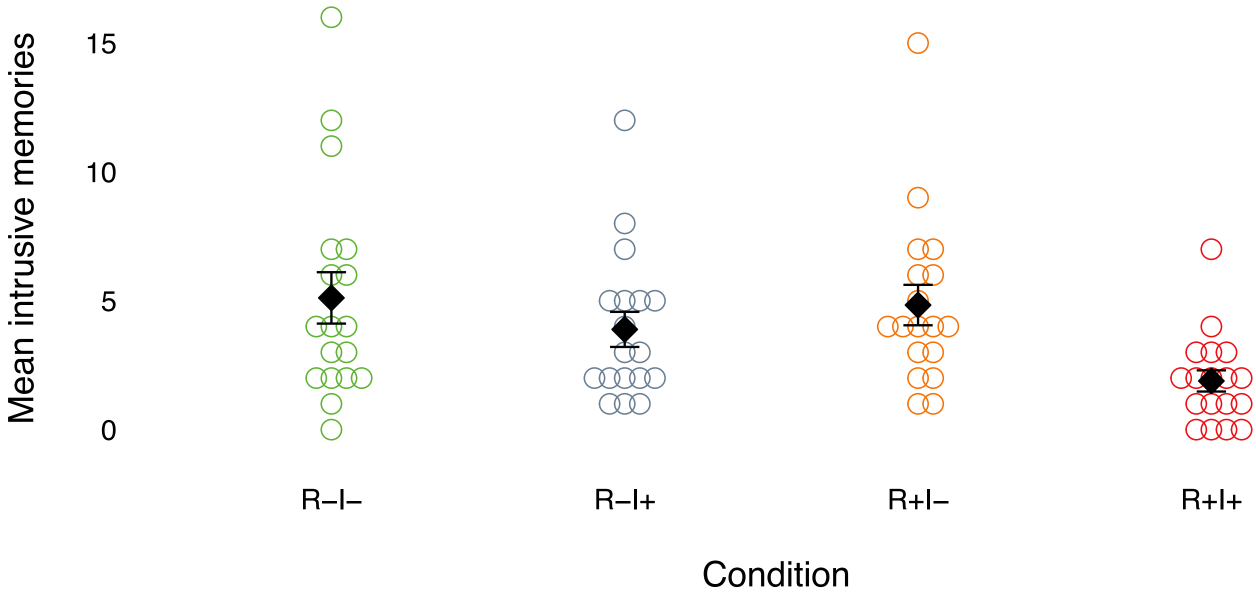


Figure 2.10: Findings of James et al. (2015, Experiment$_{\text{JH}}$ 2). Mean number of intrusive memories reported in the daily diary as a function of reminder (R) and intervention (I) condition. Black diamonds represent means, and error bars show ±SEM. Coloured circles show individual participant scores (with horizontal jitter). Plotted using data from James et al.

= .133, a significant main effect of Intervention, $F(1,68) = 7.746$, $p = .007$, and critically, no significant interaction between Reminder and Intervention, $F(1,68) = 1.323$, $p = .254$. Therefore, contrary to the conclusions of James et al. (2015), the observed effect appears to be largely driven by the independent influence of the Intervention, rather than the interaction of the Reminder and Intervention. Nevertheless, the absence of a significant interaction is difficult to interpret within the NHST framework, and it could be that a more powerful design is capable of detecting the target effect. Together with the absence of an effect on the recognition test in either experiment, and the fact that additional criteria were not assessed (C3$^?$, C8$^?$, C9$^?$, C10$^?$), the extant data obtained in this study plainly do not provide compelling support for reconsolidation theory.

2.2.3 Extinction

As discussed in subsection 1.2.3 (also see Bouton & Moody, 2004), an extinction protocol can successfully reduce responding to the CS in a conditioning paradigm, giving the impression that the previously acquired CS-US association has been 'unlearned' (Rescorla & Wagner, 1972). As a non-invasive procedure for mitigating the influence of potentially undesirable memory associations, extinction (or 'exposure therapy') is commonly used to treat disorders involving 'maladaptive memory', such as post-traumatic stress disorder (PTSD), addiction, or phobias (Bouton, 2014). However, a considerable body of evidence suggests that extinction does not involve modification of existing memory traces. Conditioned responding can typically be recovered via a number of reminder protocols (spontaneous recovery, contextual renewal, reinstatement, see subsection 1.2.3, Bouton, 2002). Consequently, the utility of extinction in clinical settings is partly undermined by the high likelihood of relapse (Bouton, 2014).

Researchers have recognised that the brief period of plasticity that is thought to occur during reconsolidation may present an opportunity for weakening or even erasing undesirable memory traces in various clinical circumstances (e.g., R. D. Lane et al., 2015; Pitman, 2011; Schwabe et al., 2014). Seeking to capitalise on this, a number of studies have examined whether performing extinction trials within the putative 'reconsolidation window' might be more effective at eliminating fear responses than standard extinction protocols (for review see Auber, Tedesco, Jones, Monfils, & Chiamulera, 2013; Kredlow, Unger, & Otto, 2015). The logic here is that reactivation of a memory trace opens a time-dependent window during which that trace can be modified. Previous studies have estimated that the upper bound for this window is 6h (i.e., the intervention becomes ineffective when delivered 6h after the reminder, Nader et al., 2000a). Therefore, reconsolidation should be disrupted when the intervention is delivered shortly after reactivation (e.g., 10m), but not when delivered after 6h.

One slightly unusual aspect of reconsolidation-extinction experiments is that they typically use a non-reinforced presentation of the CS^+ as a reminder, which is itself an extinction trial (Monfils et al., 2009; Schiller et al., 2010). The difference between a standard extinction protocol and a reconsolidation-extinction protocol is that the reminder trial is temporally isolated from the other trials. Typically, the isolated trial occurs 10m prior to the remaining extinction trials (Monfils et al., 2009; Schiller et al., 2010), but the rationale for using this interval is unclear. It is also unclear why the first trial of a standard extinction protocol should not also be considered sufficient for the initiation of reconsolidation.

Nevertheless, an initial investigation of reconsolidation-extinction in rats concluded that: "Our results are consistent with the idea that an adaptive purpose of reconsolidation is to incorporate new information at the time of retrieval, and to update a memory...in the present case leading to destabilization of the initial trace in the lateral amygdala, and the re-encoding of the once fear-inducing CS as safe" (Monfils et al., 2009, p. 955). Building on these findings, Schiller et al. (2010) sought to investigate reconsolidation-extinction in a human fear-conditioning study. During the Training Stage (Experiment$_{SP}$ 1), participants were shown two coloured squares. One square (CS^+) was paired with a mild wrist shock (US) on 38% of trials (i.e., a partial reinforcement schedule). The other square was never paired with the shock (CS^-). The purpose of the CS^- is to control for non-associative effects of the US.

Skin conductance response (SCR) was measured as an index of learning (heightened SCR is typically associated with a fear response). Approximately 24h later on Day 2, some participants received a single unreinforced presentation of the CS^+ (Reminder Stage) prior to ongoing 'extinction' trials in which both the CS^+ and CS^- were repeatedly presented without the US (Intervention Stage). Note that the reminder itself was procedurally identical to the subsequent extinction trials. The key manipulation was the time delay between the Reminder Stage and the Intervention Stage. For one group, the delay was only 10m (R^+I^+, $n = 20$). For a second group, the delay was 6h (R^+I^D, $n = 23$). For a third group, there was no reminder (R^-I^+, $n = 22$). Approximately 24h later on Day 3, all participants completed a further series of non-reinforced CS^+ and CS^- trials ('re-extinction') which also acted as a test of trace-dependent performance (Test Stage).

Fear-conditioning during the Training Stage was successful, with elevated SCRs being observed for the CS^+ relative to the CS^- in all groups. Similarly, extinction during the Intervention Stage successfully reduced this differential fear response in an equivalent manner across the three groups. The key test of reconsolidation theory was whether differential SCRs would re-emerge in the initial trials of the Day 3 Test Stage (i.e., spontaneous recovery; Rescorla, 2004). Spontaneous recovery was expected for the R^-I^+ and R^+I^D groups as the memory trace representing the CS^+ - US association should be intact. Conversely, no spontaneous recovery was predicted for the R^+I^+ group because the CS^+ - US association should have been overwritten (Schiller et al., 2010).

The spontaneous recovery effect was first examined with a two-way ANOVA with the between-subjects factor of Reminder Condition (10m, 6h, no reminder) and the within-subjects factor of re-extinction phase (early phase: trials 1-4; late phase: trials 5-8). The dependent variable was differential SCRs (CS^+ minus CS^-). A significant main effect of Reminder Condition was reported, as well as a significant Reminder*Phase interaction. However, follow-up *t*-tests were applied to a different aspect of the data. Specifically, three pairwise comparisons were conducted between the differential SCRs on the last trial of extinction (Intervention Stage) and the first trial of re-extinction (Test Stage). Significant increases were obtained for the no reminder group and 6h-delay intervention group, but not for the 10m-delay intervention group, leading the researchers to conclude that there was a spontaneous recovery effect in the former groups but not the latter. However, as the difference between significant and not significant is not necessarily itself statistically significant (Gelman & Stern, 2006; Nieuwenhuis et al., 2011), this conclusion requires a statistical comparison *between* those difference scores. In other words, it is not clear that the critical manipulations were actually effective ($C4^U$, $C5^U$, $C7^U$, $C8^U$). Other criteria were not implemented/assessed in this experiment ($C3^?$, $C6^?$, $C9^?$).

Ten to fourteen months after the Test Stage, a small subset[19] of the original participants (10m group, $n = 8$; 6h group, $n = 4$; no-reminder group, $n = 7$) returned for a reinstatement procedure (Rescorla & Heth, 1975) in order to assess propensity for recovery. Participants were given four unsignaled wrist shocks prior to a series of extinction trials (Recovery Test). The researchers reported a 'reinstatement index' which in the main text is defined as the "difference in the conditioned fear response at the end of re-extinction after the initial spontaneous recovery test and the conditioned fear response immediately after reinstatement

[19] Some additional participants who returned for the Recovery Test ($n = 4$) were later excluded because they 'failed to re-extinguish after the spontaneous recovery test' or 'showed no measurable responses to the shocks during reinstatement'. Whether these exclusions had an appreciable influence on the outcome of the main analysis is not reported.

1 year later" (Schiller et al., 2010, p. 51). However, in the supplementary materials (p.3) it is noted that "The first CS+ and CS- trial were disregarded due to the orienting response at the beginning of the session". Other researchers have pointed out that this approach may be problematic because the first trial, being an extinction trial itself, will likely have attenuated the fear response, potentially masking recovery effects (Kindt & Soeter, 2013; Soeter & Kindt, 2011).

For analysis of recovery effects, the researchers did conduct a statistical comparison of the time-point differences between groups (cf. above). A two-way ANOVA with a between-subjects factor of Reminder Group (10m, 6h, no reminder), a within-subjects factor of stage (Test Stage, Recovery Test), and differential SCR score (CS^{+} minus CS^{-}) as the dependent variable, revealed no significant interaction ($p = .07$). Nevertheless, a one-tailed t-test suggested that the reinstatement index (see above) was significantly larger in the 6h/no-reminder groups combined compared to the 10m group ($t = 1.75$, $p = .049$[20]). Based only on marginally significant differences and small sample sizes, it seems reasonable to remain cautious about the robustness of these effects ($C10^{U}$). Furthermore, using the reported t value it is possible to calculate a Bayes Factor ("default" JZS prior, Cauchy distribution with scale $r = .707$, Rouder, Speckman, Sun, Morey, & Iverson, 2009) in order to shed light on the evidentiary support for the (directional) alternative hypothesis (H_1: 6h/no-reminder groups - 10m group > 0) relative to the null hypothesis (H_0: 6h/no-reminder groups - 10m group = 0). The Bayes Factor indicates only weak evidence in favour of H_1 relative to H_0 ($BF_{10} = 2.06$).

In a second experiment ($\text{Experiment}_{\text{SP}}$ 2), Schiller et al. (2010) employed a similar procedure to $\text{Experiment}_{\text{SP}}$ 1 ($C1^{Y}$, $C2^{Y}$) with some key modifications. Firstly, a within-subjects (rather than between-subjects) reminder/no-reminder manipulation was employed ($n = 18$). During the Training Stage, there were three square stimuli: two were associated with wrist shock on 38% of trials (CSa^{+} and CSb^{+}) and one was never associated with shock (CS^{-}). The Day 2 Reminder Stage consisted of a single presentation of CSa^{+} and CS^{-}. CSb^{+} was not presented. The Intervention Stage was conducted 10m later. The Final Test on Day 3 employed a

[20]Note that in the original article the p-value is reported as $p < .05$, but the precise value ($p = .049$) can be recalculated from the t value.

reinstatement protocol (see above).

In order to examine the recovery of fear, the researchers employed a two-way ANOVA with CS-type (CSa^+, CSb^+ and CS^-) and re-extinction phase (early phase: trials 1-4; late phase: trials 5-8) as independent variables, and SCRs as the dependent variable. A significant CS-type*phase interaction is reported, but as above, the follow-up *t*-tests were applied to a different aspect of the data. Specifically, SCRs were compared between the last trial of the Intervention Stage and the first extinction trial of the Test Stage (i.e., post-reinstatement). A series of *t* tests indicated significant recovery only for CSb^+ (i.e., the R-I+ condition), and not for CSa^+ or CS^-. As outlined above, it is important to examine whether these effects are significantly different from each other in order to establish that the critical reminder manipulation has been effective. As the data are not publicly available, it is not possible to assess whether this is the case or not ($C4^U$, $C5^U$, $C6^U$ $C7^U$). Other criteria were not implemented/assessed ($C3^?$, $C5^?$, $C8^?$, $C9^?$, $C10^?$).

Similar issues are also apparent in other extinction studies that are interpreted as yielding evidence favourable to reconsolidation theory (Agren et al., 2012; Björkstrand et al., 2015). In Agren et al. (2012), following fear-conditioning on Day 1 (Training Stage), all participants completed a Reminder Stage 24h later on Day 2 ($C1^Y$). Subsequently, participants performed extinction trials (Intervention Stage) either after a 10m delay (R^+I^+ group) or after a 6h delay ($R+I^D$; $C2^Y$). On Day 3, all participants had a functional magnetic resonance imaging scan but no behavioural measures were recorded. Finally, on Day 5, all participants completed a reinstatement protocol. Return of fear (ROF) was operationalised as SCRs for the CS^+ on the first post-reinstatement trial of Day 5 minus CS^+ SCRs on the final extinction trial of Day 2. ROF scores are shown in Figure 2.11. Two one-sample *t*-tests were conducted on the ROF scores for each group individually, and a statistically significant effect was observed in the 6h group, $t(10) = 2.72$, $p = 0.02$, but not in the 10m group, $t(8) = 0.23$, $p = 0.82$. The researchers concluded that "extinction training initiated during reconsolidation abolishes fear expression by erasing a memory trace in the amygdala" (Agren et al., 2012, p. 1552).

The analysis used to support this conclusion is problematic for two primary reasons. Firstly, it does not account for the CS- which is supposed to act as a control for the non-associative

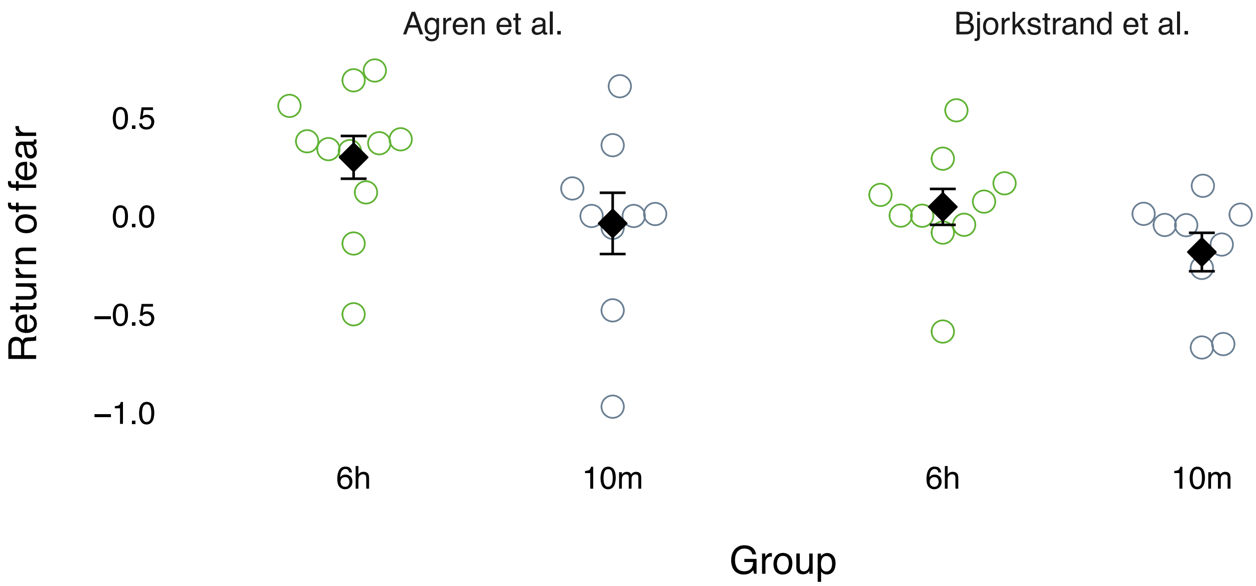


Figure 2.11: Return of fear (skin-conductance response; see main text for precise operationalisation) in Agren et al. (2012) and Bjorkstrand et al. (2015). Black diamonds represent means, and error bars show ±SEM. Coloured circles show individual participant scores (with horizontal jitter). Plotted using data from Agren et al. and Bjorkstrand et al. (see main text for details).

effects of the US. Secondly, it relies on a statistically significant effect in one group (6h), but not in the other (10m), without testing if they are significantly different from each other (Gelman & Stern, 2006; Nieuwenhuis et al., 2011). Fortunately, a subset of the data from this study was available (Agren, personal communication, March 7, 2016) and it was possible to run an alternative analysis. A three-way mixed ANOVA with the factors Group (10m, 6h), CS (plus, minus), and Stage (last trial of extinction, first trial post-reinstatement) revealed no statistically significant main effects or interactions (Table 2.3).

Furthermore, a Welch two-sample t-test comparing the CS^+ only ROF scores (see above) *between* the 6h and 10m groups also indicated no significant effect, $t(14.87) = -1.77$, $p = .097$ (see Figure 2.11). A Bayes Factor ("default" JZS prior, Cauchy distribution with scale $r =$ .707, Rouder et al., 2009) can also be calculated to quantify the evidentiary support for the alternative hypothesis (H_1: 10m group - 6h group $\neq 0$) relative to the null hypothesis (H_0: 10m group - 6h group $= 0$). The Bayes Factor indicates that the data do not adequately differentiate between H_1 and H_0 (i.e., the data are inconclusive, $BF_{10} = 1.14$).

Table 2.3: ANOVA table for return of fear re-analysis based on Agren et al. (2012a).

Term	F	df_1	df_2	p	η^2_G
Group	0.96	1	18	.339	.01
CS	2.16	1	18	.159	.03
Group × CS	0.01	1	18	.930	.00
Group × Stage	3.97	1	18	.062	.07
CS × Stage	0.87	1	18	.364	.01
Group × CS × Stage	0.16	1	18	.694	.00

Group, 10m/6h reminder-intervention delay. *CS* = plus/minus. *Stage* = last trial of Day 2 extinction/first trial of Day 5 post-reinstatement.

In an 18-month follow-up experiment, Björkstrand et al. (2015) were able to recall 19 out of the 22 original participants and conducted an additional ROF test. In this case, recovery was examined in a 4-trial reacquisition session during which the CS^+ was again paired with the US. Mean SCRs across these four trials were compared between the CS^+ and the CS^- independently within each group (10m, 6h). A significant effect (reacquisition) was reported in the 6h group, $t(9) = 2.08$, $p = .034$, but not in the 10m group, $t(8) = 0.95$, $p = .185$ (one-tailed tests). In addition, ROF scores were calculated (see Figure 2.11), operationalised as CS^+ SCRs on the last trial of extinction on Day 2 relative to mean SCRs over 4 trials of reacquisition 18-month later (in contrast to Agren et al., where only the first post-reinstatement trial was used). Analysis of this ROF score did involve a statistical comparison of the difference scores (cf. Agren et al.), and indicated a marginally significant effect, $t(17) = 1.72$; p = .052 (one-tailed). The conclusion was that "disrupting reconsolidation have (*sic*) long lasting behavioral effects and may permanently erase the fear component of an amygdala-dependent memory" (Björkstrand et al., 2015, p. 1). However, as was the case for Agren et al., an alternative analysis of the data does not appear to support this conclusion.

Table 2.4: ANOVA table for reacquisition re-analysis based on Bjorkstrand et al. (2015).

Term	F	df_1	df_2	p	η^2_G
Group	2.50	1	17	.132	.07
CS	5.19	1	17	.036	.12
Group × CS	1.55	1	17	.230	.04

Group, 10m/6h reminder-intervention delay. *CS* = plus/minus.

Firstly, the reacquisition analysis does not involve a statistical comparison between the two groups (10m, 6h) which is problematic for the reasons outlined above. A re-analysis of the data (which is publicly available here: http://dx.doi.org/10.5061/dryad.v1d06/) using a two-way mixed ANOVA with Group (10m, 6h) as a between-subjects factor, and CS (plus, minus) as a within-subjects factor, indicates only a significant main effect of CS (see Table 2.4). The absence of a group by CS-type interaction undermines the claim that there was reacquisition in the 6h group but not in the 10m group (Björkstrand et al., 2015).

Moreover, the ROF analysis is problematic firstly because it does not include the CS^-. Unfortunately, insufficient information is included in the available datasets to conduct a re-analysis including the CS^-. The second problem pertains to the evidential value of the data, which were only marginally significant. A Bayes Factor (“default” JZS prior, Cauchy distribution with scale $r = .707$, Rouder et al., 2009) quantifying the evidentiary support for the (directional, to correspond to the use of a one-tailed t-test) alternative hypothesis (H_1: 10m group - 6h group < 0) relative to the null hypothesis (H_0: 10m group - 6h group $= 0$). The Bayes Factor indicates that the data do not adequately differentiate between H_1 and H_0 ($BF_{10} = 1.08$).

In summary, the findings of Schiller et al. (2010), Agren et al. (2012), and Björkstrand et al. (2015) appear to be numerically consistent with several key aspects of reconsolidation theory, but the robustness of the effects is uncertain, and Bayesian analysis indicates that the evidentiary value is extremely weak[21]. A number of subsequent studies have attempted to replicate the findings with varying degrees of success. Whilst some researchers have reported successful reconsolidation-extinction effects (D. C. Johnson & Casey, 2015; Oyarzún et al., 2012; Schiller, Kanen, LeDoux, Monfils, & Phelps, 2013; Steinfurth et al., 2014; Y. X. Xue et al., 2012), others have found the critical retrieval manipulation to be ineffective (Fricchione et al., 2016; Golkar et al., 2012; Kindt & Soeter, 2013; Klucken et al., 2016; Meir Drexler et al., 2014; Soeter & Kindt, 2011; Warren et al., 2014). Attempts to replicate the original demonstration of reconsolidation-extinction in rats (Monfils et al., 2009) have also resulted in mixed outcomes (cf. Baker, McNally, & Richardson, 2013; W. Y. M. Chan, Leung, Westbrook, & McNally, 2010; Costanzi, Cannas, Saraulli, Rossi-Arnaud, & Cestari, 2011; Ishii et al., 2015; Jones, Ringuet, & Monfils, 2013). Because a number of parameters vary between these replications, it can be difficult to ascertain whether a putative reconsolidation process is being modulated by subtle theoretically-irrelevant procedural differences, or theoretically-relevant 'boundary conditions' (Auber et al., 2013; Haaker et al., 2014). These mixed findings could also indicate that extinction, and more generally new learning, is not capable of disrupting reconsolidation processes. It could also be the case that there is no reconsolidation process occurring at all.

2.3 Theoretical quagmire

2.3.1 Stuck in the mud: Boundary conditions and theoretical progress

Since its (re)birth a decade and a half ago (Nader et al., 2000a; Przybyslawski & Sara, 1997; Sara, 2000a), the theory of reconsolidation has been buoyed by considerable empirical attention and theoretical excitement (Alberini, 2011; Besnard et al., 2012; Dudai, 2004; Nadel & Land, 2000; Nader et al., 2000b; Nader & Hardt, 2009; Nader, 2015; Schwabe et al.,

[21] Note that Agren et al. and Bjorkstrand et al. also reported correlations between blood oxygen level dependent (BOLD) activity in the amygdala and the behavioural return of fear measure. However, we have focused on whether there is a behavioural effect consistent with reconsolidation theory. It is unclear how much can be inferred from brain imaging when the behavioural data indicates no effect of the critical manipulation (Henson, 2005; Wilkinson & Halligan, 2004).

2014; Tronson & Taylor, 2007). Conjecture that reconsolidation may be a memory updating mechanism furnishing traces with a capacity for both persistence and plasticity is widespread (e.g., Dudai, 2009; Exton-McGuinness et al., 2015; Hardt et al., 2010; J. L. Lee, 2009; Sara, 2000b). However, in many respects the present reconsolidation literature is recapitulating the same issues that dogged its forebear, consolidation, during the previous century (R. R. Miller & Matzel, 2000, 2006; Millin, Moody, & Riccio, 2001; Riccio, Moody, & Millin, 2002; Riccio et al., 2006; Sara & Hars, 2006).

The assumption that the stabilisation of a newly acquired trace renders it relatively immune to subsequent interference attempts (see Figure 1.10) was challenged decades ago by studies showing that a retrieval cue could restore vulnerability to various pharmacological and electroconvulsive interventions (see paragraph 1.2.7.8.3; e.g., Lewis et al., 1972; Misanin et al., 1968; Richardson et al., 1982). However, rather than attempting to reconcile these findings within the prevailing zeitgeist (cf. Nader et al., 2000a), many early theorists viewed these *cue-dependent amnesia* (CDA) effects as just one of a number of sharp thorns in the side of consolidation theory (Lewis & Maher, 1965; R. R. Miller & Springer, 1973; Riccio & Richardson, 1984; Spear, 1973). Perhaps the most problematic prediction of consolidation theory was that experimentally-induced amnesia would be permanent (because the underlying impairment was storage-based). Contrary to this notion, a considerable number of investigations reported that amnesia effects were only transient (suggesting the impairment was retrieval-based, R. R. Miller & Springer, 1973; Sara & Hars, 2006). Similarly, several early CDA studies also found that trace-dependent performance could at least partially recover, either spontaneously or via a reinstatement protocol (e.g., Bregman et al., 1976; Judge & Quartermain, 1982; Mactutus et al., 1982).

Many contemporary investigations of CDA/reconsolidation also report that amnesia effects are transient (e.g., Anokhin & Tiunova, 2002; Eisenberg & Dudai, 2004; Gisquet-Verrier et al., 2015; Lattal & Abel, 2004; Trent et al., 2015). This is highly problematic for a theory which suggests that amnesic agents act to disrupt restabilisation of memory traces, rendering them irreversibly impaired or even erased (Nader, 2003a; Nader & Hardt, 2009). Nevertheless, as was the case with consolidation (see subsubsection 1.2.7.8), investigation of recovery effects has yielded a disconcerting quantity of conflicting evidence. Several studies report that amnesia effects in reconsolidation studies are lasting, despite attempts at reinstatement (e.g., Boccia, Blake, Acosta, & Baratti, 2005; Debiec et al., 2002; Duvarci & Nader, 2004). In contrast, several studies have not been able to obtain cue-dependent amnesia effects at all (Biedenkapp & Rudy, 2004; Blaiss & Janak, 2007; Cammarota, Bevilaqua, Medina, & Izquierdo, 2004; Font & Cunningham, 2012; Taubenfeld, Milekic, & Monti, 2001), emulating earlier replication difficulties (see paragraph 1.2.7.8.3, Banker et al., 1969; Dawson & McGaugh, 1969; Gold & King, 1972; Jamieson & Albert, 1970; Squire et al., 1976).

The outcomes of our systematic review (section 2.1) and selective review (section 2.2) also reveal that there is a considerable quantity of evidence inconsistent with the theory, and inconsistent with other evidence. Many studies include multiple cases, some of which appear to yield reconsolidation effects, and some of which do not (see Figure 2.7). Furthermore, many prominent findings have proved difficult to obtain in subsequent replication attempts (e.g., cf. Hardwicke et al., 2016; Walker et al., 2003; and cf. Kindt & Soeter, 2013; Schiller et al., 2010). The typical strategy used to accommodate situations such as these is to propose that a 'boundary condition' of the theory has been identified (e.g., J. L. Lee, 2009; Nader & Hardt, 2009; Tronson & Taylor, 2007; Walker & Stickgold, 2016). Proposed boundary conditions include length of reactivation (Eisenberg et al., 2003), prediction error (Exton-McGuinness et al., 2015), trace strength (Suzuki, 2004), reminder context (Hupbach et al., 2008), participant age (St Jacques et al., 2015), session time (Walker & Stickgold, 2016), valence of learning materials (Schwabe & Wolf, 2010), valence of intervention (Strange et al., 2010), sleep/wake state (Diekelmann et al., 2011), and memory age (Coccoz et al., 2013). Unfortunately, many of the boundary conditions are themselves based on conflicting evidence. For instance, it has been proposed that exposure to the training context is both sufficient and necessary to trigger reconsolidation (Hupbach et al., 2008; also see DeVietti & Holliday, 1972), but other studies indicate that it is neither necessary (Kroes et al., 2014), nor sufficient (Forcato et al., 2009).

It has also been proposed that older memory traces do not undergo reconsolidation, leading some theorists to suggest that cue-dependent amnesia effects actually represent disruption of a lingering consolidation process (Alberini, 2011; Dudai & Eisenberg, 2004). In apparent support of this notion, a number of studies have reported reconsolidation effects in 'younger', but not 'older' memory traces. For instance, reconsolidation effects have been observed with training-reminder intervals at 1d but not 8d (Forcato et al., 2013), 6d but not 20d (Coccoz et al., 2013), 2d and 7d, but not 14d or 28d (Milekic & Alberini, 2002), and 1d and 28d, but not 7d (Wichert et al., 2011). However, other studies have observed reconsolidation effects for 'older' memory traces, with training-reminder intervals of 7d (Steinfurth et al., 2014), 14d (Nader et al., 2000a), 30d (Einarsson & Nader, 2012), and 60d (Dębiec & LeDoux, 2004). A boundary condition of memory age also seems incompatible with the notion that reconsolidation routinely updates traces over their lifespan (J. L. Lee, 2009; Exton-McGuinness et al., 2015).

An additional and much discussed candidate boundary condition is prediction error (Exton-McGuinness et al., 2015; J. L. Lee, 2009). According to this idea, reconsolidation will only be triggered when a prediction error signal occurs at the reminder stage. The idea is that trace updating is only adaptive when some change in the environment necessitates a concomitant change in the information content of the memory system (J. L. Lee, 2009). A number of studies appear to indicate a role for prediction error because reconsolidation effects are only obtained when there is an apparent 'mismatch' between what the experimental subject should expect, and what actually happens (Forcato et al., 2009; Pedreira et al., 2004; Morris et al., 2006; Rodriguez-Ortiz, 2005; Sevenster, Beckers, & Kindt, 2014). However, many non-human animal studies have employed conditioning protocols in which a reinforced presentation of the previously learned contingency (CS^+ and US) is used to reactivate the underlying trace at the Reminder Stage. Therefore, the expectations of the subject are fulfilled, and there is presumably no prediction error. Nonetheless, these studies still report reconsolidation effects (Duvarci & Nader, 2004; Eisenberg & Dudai, 2004; Sangha et al., 2003). Similarly, several apparent demonstrations of reconsolidation in humans involve specific-active or specific-passive reminders (see Figure 2.5, and subsubsection 2.1.13.1) in which the participant is tested on the knowledge they acquired during training (e.g., J. C. K. Chan & LaPaglia, 2013; Beukelaar et al., 2014; Walker et al., 2003; Wichert et al., 2011), or re-presented with aspects of the learning material (e.g., James et al., 2015; St Jacques & Schacter, 2013; Wirkner et al., 2015). It seems unlikely that these protocols involve prediction error.

The numerous appeals to boundary conditions is indicative of a Lakatosian defense intended to preserve the 'core' of the theory by adjusting its auxiliary assumptions in instances where its predictions have not been met (Lakatos, 1978; Meehl, 1990a). Such a defense may be warranted when a theory has a 'good track record', i.e., it can successfully account for a wide range of other relevant phenomena. This is arguably the case for the consolidation/reconsolidation framework as it offers a reasonable explanation for the widespread evidence of temporally-graded amnesia observed across multiple species and paradigms (see Chapter 1). Informally, it may seem like "we are on to something" and it would certainly be rash to discard a successful theory at the merest hint of falsification. The boundary condition defense therefore seems like a rational strategy in principle.

The strategy can become problematic however, when it is only deployed in the context of individual empirical reports, and its implications are not considered at the field level. For instance, a new *ad hoc* theoretical assumption might appear to provide an adequate explanation of a discrepancy between data and theory in the case of a non-replication (cf. Hardwicke & Shanks, 2016; Walker & Stickgold, 2016). However, calibrating a theory to a specific dataset on an *ad hoc* basis may only explain the idiosyncratic scenario under scrutiny, whilst reducing the overall explanatory coverage of the theory at the field level (i.e., 'over-fitting'). Boundary conditions should not only be invoked in cases of theoretical failure, but must also be applied to cases of theoretical success, otherwise they are not really boundary conditions at all, but new 'sub-theories'. In other words, boundary conditions should not just be used to 'explain away' individual cases that are inconsistent with the theory, but should be integrated with existing theoretical tenets, and applied to the existing evidence base. The danger is that it is easy to 'over-fit' an explanation once the data have been observed (Kerr, 1998; Wagenmakers, Wetzels, Borsboom, Maas, & Kievit, 2012). Boundary conditions are only likely to be useful theoretical tools if they also accommodate cases that have previously offered support for the theory (i.e., the explanatory coverage of the theory is maintained), and generate new predictions that can be verified empirically (cf. Hardwicke & Shanks, 2016; Walker & Stickgold, 2016). The proliferation of unverified boundary conditions that are incompatible with aspects of the extant data, and do not generate precise testable predictions, has led to theoretical confusion rather than theoretical progress. From a Lakatosian perspective, the research programme is degenerating (Lakatos, 1978).

2.3.2 Recommendations to foster theoretical progress

How can reconsolidation extract itself from this theoretical quagmire? When one is stuck in the mud, conventional wisdom suggests “Don’t panic!” (Adams, 1979). A number of other strategies, based on the principles of open, cumulative, and collaborative science may also be useful in our specific context.

2.3.2.1 Direct replication

Years of intensive empirical investigation (Figure 1.12) have yielded a disappointing bounty of ambiguous and conflicting evidence and it has become extremely difficult to isolate the conditions under which reconsolidation effects will reliably emerge. A possible reason for this is the absence of direct replication (Simons, 2014). In a direct replication (e.g., Hardwicke et al., 2016), the idea is to repeat the methodology of a previous study as closely as possible. The goal is to identify whether the previously reported effect is reliable. This is important because the findings of an individual study can be entirely spurious for a number of reasons: “The undetected equipment failure, the rare and possibly random human errors of procedure, observation, recording, computation, or report are known well enough to make scientists wary of the unreplicated experiment. When we add the possibility of the random ‘fluke’ common to all sciences, the fact of individual organismic differences, and the possibility of systematic experimenter effects, the importance of replication looms larger still to the behavioral scientist” (Rosenthal, 1990, p. 2). The dramatic prevalence of low-sample size studies (see Figure 2.6) should make us especially skeptical of the reliability of the field’s major empirical claims (Button et al., 2013; Meehl, 1990b).

Direct replications are rarely reported in the field of reconsolidation (cf. Hardwicke et al., 2016), although arguably ‘conceptual’ replications are more common, particularly in studies based on the well-established fear-conditioning and extinction paradigm (Golkar et al., 2012;

Kindt & Soeter, 2013; Soeter & Kindt, 2011). In a conceptual replication, one follows the methods of a previous study to some extent, but key parameters are varied in order to test the generalisability of the theory (Schmidt, 2009). Conceptual replications are most useful when part of a 'replication battery' that systematically and explicitly manipulates key theoretically-relevant parameters on a continuum of similarity based on the original study (Rosenthal, 1990; Schmidt, 2009). The problem with a conceptual replication that differs haphazardly, rather than systematically, from the original study, is that the source of any deviation from the original study's findings is difficult to identify.

For example, a post-retrieval extinction study conducted by Soeter and Kindt (2011; also see Kindt & Soeter, 2013) was unable to replicate a reconsolidation effect reported in a similar study conducted by Schiller et al. (2010). The absence of a reconsolidation effect in the replication studies is an important finding because the predictions of reconsolidation theory were found wanting. However, how can we explain the discrepancy with the original finding? Although reasonably similar, there were several differences between the studies. For example, Soeter and Kindt (2011) used fear-relevant stimuli (pictures of spiders), whereas Schiller et al. (2010) used fear-irrelevant stimuli (coloured squares). Could it be the case that the reconsolidation effect is contingent on the use of fear-irrelevant stimuli? Or perhaps the original finding was a false-positive? It is difficult to draw strong conclusions because the replication also varied in several other ways from the original study (e.g., different reinforcement schedule). Indeed, the ambiguity of a 'failed' conceptual replication may increase the likelihood that it ends up in the file-drawer (Fanelli, 2012; Rosenthal, 1979), whilst successful conceptual replications are readily published because they demonstrate positive evidence for the theory (Pashler & Harris, 2012; Simons, 2014). A more fruitful approach therefore would be to combine the conceptual replication with a direct replication, and manipulate only the key parameter of interest. If the direct replication obtains a similar effect to the original, and the conceptual replication with (say) fear-relevant stimuli does not, then one can be more certain that the nature of the stimuli is in some way moderating the effect. It could also be that the direct replication itself does not obtain a similar finding to the original (e.g., Hardwicke et al., 2016), in which case, depending on the evidential value of the replication (i.e., sample size etc.) relative to the original study, the reliability of the original finding should be placed in doubt (see Chapter 3).

Widespread methodological heterogeneity creates considerable problems for those attempting to summarise the state of the field. In a review of the reconsolidation-extinction literature, Auber and colleagues (Auber et al., 2013) diligently attempted to extract the relevant boundary conditions or moderators that could be causing discrepant findings (also see Kredlow et al., 2015). However, they were eventually forced to note that “seemingly slight differences in methodologies can yield drastically varied results” (Auber et al., 2013, p. 633). Fostering a healthy culture of direct replication can help to identify which of these methodological differences are actually important, and which simply reflect the poor reliability of one or more of the studies under scrutiny (Ioannidis, 2012, 2014).

2.3.2.2 Open data

Published studies rarely upload their raw data to a third-party repository so that it is publicly available for other scientists to access (e.g., Kidwell et al., 2016). Additionally, despite the data being supposedly available on request according to the standards of professional organisations and journals, this is rarely the case in practice (Vanpaemel, Vermorgen, Deriemaecker, & Storms, 2015; Wicherts, Borsboom, Kats, & Molenaar, 2006). Data availability is important because published articles merely represent the authors’ summary account of the study, and not the evidence itself, which is the data. Given that scientists are only human (Hardwicke, Jameel, Jones, Walczak, & Magis-Weinberg, 2014), and operate within scientific ecosystems characterised by unscientific values (M Bakker, Dijk, & Wicherts, 2012; Giner-Sorolla, 2012; Nosek, Spies, & Motyl, 2012), their interpretations of the data are highly susceptible to bias and error, even when this is unintentional (Marjan Bakker & Wicherts, 2011; Chamberlin, 1965; John, Loewenstein, & Prelec, 2012; O’Boyle, Banks, & Gonzalez-Mule, 2014).

A powerful defense against fooling ourselves and others (Feynman, 1974), is to make data available to the scientific commons so it can be checked, re-analysed, and incorporated into quantitative summaries via meta-analysis (Ioannidis, 2014). Concrete examples are available in section 2.2, where re-analysis of data from published studies suggests that critical interactions are not statistically significant, and the data are relatively inconclusive in determining between the null and reconsolidation hypotheses.

2.3.2.3 Pre-registration

The rapid growth of the reconsolidation field (see Figure 1.12) has been accompanied by the rapid proliferation of *ad hoc* theoretical adjustments, or 'boundary conditions'. As outlined above, invoking a boundary conditions argument can be a useful tool for theoretical advancement, but only when there is a clear distinction between exploratory 'hypothesis-generating' and confirmatory 'hypothesis-testing' phases of scientific inquiry (Kerr, 1998; Wagenmakers et al., 2012; also cf. Hardwicke & Shanks, 2016; Walker & Stickgold, 2016). When considering the various components of a study (hypotheses, methods, analysis protocols), pre-registration simply makes clear what was planned, and what was not. As well as constraining untoward or unintended exploitation of various researcher degrees of freedom (Gelman & Loken, 2014; Simmons, Nelson, & Simonsohn, 2011), pre-registration makes transparent whether a proposed boundary condition is based on an *a priori* hypothesis, or *post hoc* conjecture (Kerr, 1998; Wagenmakers et al., 2012). Both situations are perfectly valid aspects of scientific inquiry, but only when the distinction is made explicit. If a boundary condition is proposed *a priori*, and confirmed by experiment, the scientist (and others) can be more confident that it is genuine. Conversely, when a boundary condition is proposed *post hoc* in order to accommodate anomalous data within a particular theoretical framework, the scientist (and others) might be less convinced and await empirical confirmation before drawing strong conclusions (Hardwicke & Shanks, 2016; Wagenmakers et al., 2012).

2.3.2.4 Collaborative research

A general recommendation that has numerous associated benefits, is to encourage more large-scale, collaborative research in which multiple independent laboratories work together towards a shared goal (Ioannidis, 2014; Klein et al., 2014; Open Science Collaboration, 2015; Simons, Holcombe, & Spellman, 2014). In collaborative research, scientists can pool otherwise sparse resources. For example, by distributing responsibility for data collection across multiple teams, it is possible to recruit much larger sample sizes than are typically obtained by a single team working alone (see Figure 2.6). Recruiting large samples will lead to more precise effect size estimates (Lakens & Evers, 2014) and help to improve the reliability of the field (Button et al., 2013). Large-scale collaboration also fosters other open practices (open data, open analysis code, open materials, pre-registration) which naturally support efficient collaboration (Ioannidis, 2014; Simons et al., 2014).

2.4 Summary and conclusions

The outcomes of our systematic review (section 2.1) and selective review (section 2.2) suggest that the theory of reconsolidation-mediated memory updating lacks a compelling evidence-base. This is partly due to the considerable ambiguity surrounding the large number of cases where key criteria have not been implemented or assessed (see Figures 2.2 and 2.8). Initially promising findings have proved difficult to replicate (e.g., cf. Hardwicke et al., 2016; Walker et al., 2003; and cf. Kindt & Soeter, 2013; Schiller et al., 2010), or can be explained without recourse to reconsolidation theory (e.g., Hupbach et al., 2007; Sederberg et al., 2011; and cf. J. C. K. Chan & LaPaglia, 2013, Chapter 4). Several studies that are considered to demonstrate convincing reconsolidation effects (Agren et al., 2012; Björkstrand et al., 2015; James et al., 2015; Schiller et al., 2010), no longer seem compelling when re-analysis indicates that critical interactions are absent, and their evidentiary support for the reconsolidation hypothesis is weak relative to the null (see section 2.2). Although the field is often presented as documenting a weighty catalogue of theoretical successes, it is in fact imbued with numerous theoretical failures (see Figure 2.7). Reconciling these instances with the central tenets of the theory is far from straightforward due to considerable methodological heterogeneity coupled with a lack of systematic, direct replications. This has resulted in frequent appeals to ‘boundary conditions’ recruited to protect the core tenets of the theory. However, even these boundary conditions are based on bodies of conflicting evidence, resulting in a highly complex theoretical landscape in which various sub-versions of the theory are distributed across the field and called upon as required. As a result, the field of reconsolidation has stumbled into a quagmire, and the central tenets of the theory are in doubt.

Table 2.5: Systematic review of human reconsolidation literature (behavioural interventions). **DV** = dependent variable. **R** = reminder type (*S* = Specific, *G* = General, *A* = Active, *P* = Passive). **Effect** = Claimed reconsolidation effect (*D* = Destructive, *C* = Constructive, *N* = No reconsolidation effect). **N** = sample size of critical R^+I^+ group. **C** = Criterion (see main text). *:* = Not applicable. ˆ = As above.

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Walker et al. (2003)	motor sequence learning	:	accuracy	new learning	S-A	D	16	Y	Y	Y	Y	Y	?	?	?	Y	?
	ˆ	:	speed	ˆ	ˆ	N	ˆ	Y	Y	N	Y	Y	?	?	:	:	:
Hupbach et al. (2007)	list learning	:	correct (free recall)	new learning	G-A	N	12	Y	Y	?	N	?	Y	?	:	:	:
	ˆ	:	intrusions (free recall)	ˆ	ˆ	C	ˆ	Y	Y	?	Y	?	Y	?	?	Y	?
Forcato et al. (2007)	paired-associates	E1: Test L1-L2	L1 errors (cued-recall)	new learning	S-A	D	10	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E1: Test L2-L1	L1 errors (cued-recall)	new learning	S-A	D	10	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E2: Test L1-L2	L1 errors (cued-recall)	new learning	S-A	D	10	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E2: Test L2-L1	L1 errors (cued-recall)	new learning	S-A	D	10	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E3: Test L1-L2	L1 errors (cued-recall)	new learning	S-A	D	10	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E3: Test L2-L1	L1 errors (cued-recall)	new learning	S-A	D	10	Y	Y	?	?	Y	?	?	?	?	?

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	paired-associates	Fig 4 E2[a]	L1 errors (cued-recall)	new learning	S-A	D	10	Y	Y	?	?	Y	?	?	?	?	?
Hupbach et al. (2008)	list learning	:	correct (free recall)	new learning	G-P	N	12	Y	Y	?	N	?	?	?	:	:	:
	list learning	:	intrusions (free recall)	⌢	⌢	C	⌢	Y	Y	?	?	?	?	?	?	?	?
Schwabe & Wolf (2009)	autobio-graphical narratives	neutral-valence	number of 'details' (cued recall)	new learning	S-A	D	12	Y	Y	?	Y	Y	Y	Y	?	?	?
	autobio-graphical narratives	negative-valence	number of 'details' (cued recall)	new learning	S-A	N	⌢	Y	Y	?	N	N	N	N	?	?	?
	autobio-graphical narratives	positive-valence	number of 'details' (cued recall)	new learning	S-A	N	⌢	Y	Y	?	N	N	N	N	?	?	?
Chan et al. (2009)	post-event misinforma-tion	:	recall probability correct (MMFR)	new learning	S-A	D	24	N	Y	?	U	U	U	U[b]	?	?	?

[a]This is the second of two 'Experiment 2s' that are reported in this article - the numbering scheme is on a per figure basis.

[b]there were no significant difference between recall in the R^+I^+, R^-I^-, and R^-I^+ groups, but performance was lower than in the R^+I^- group (with a significant R*I interaction). The authors argue that the presence of a testing effect in the no-interference groups ($R^+I^- > R^-I^-$) and absence of testing effect in the interference groups ($R^+I^+ = R^-I^+$) is consistent with reconsolidation theory (i.e., the testing effect may have masked a reconsolidation effect). However, in a subsequent study (Chan et al. 2011) a similar pattern was observed even when the intervention was delayed suggesting that the effect cannot be attributed to disruption of reconsolidation (see note accompanying Chan et al. 2011).

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Forcato et al. (2009)	paired-associates	Fig 2b[c]	L1 errors (cued-recall)	new learning	S-A	D	13	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	Fig 3a	L1 errors (cued-recall)	new learning	G-P	N	13	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	Fig 3b	L1 errors (cued-recall)	new learning	S-A	N	13	Y	Y	?	?	Y	?	?	?	?	?
Hupbach et al. (2009)	list learning	:	correct (old/new recognition)	new learning	G-P	N	10	Y	Y	?	N	?	?	?	:	:	:
	^	:	correct (source attributions)	^	^	C	^	Y	Y	?	Y	?	?	?	?	?	?
Zhao et al (2009)	list learning	positive words	correct (free recall)	stress	S-A	D	21	Y	Y	?	?	Y	?	?	?	?	?
	list learning	negative words	correct (free recall)	stress	S-A	D	^	Y	Y	?	?	Y	?	?	?	?	?
	list learning	neutral words	correct (free recall)	stress	S-A	N	^	Y	Y	?	?	N	?	?	?	?	?
Schwabe & Wolf (2010)	autobio-graphical narratives	neutral-valence event	number of 'details' (cued recall)	stress	S-A	D	16	Y	Y	Y	Y	Y	Y	U[d]	?	?	?
	autobio-graphical narratives	negative-valence event	number of 'details' (cued recall)	stress	S-A	N	^	Y	Y	N	N	N	N	U[d]	?	?	?

[c] Experiments are not numbered in this article so they will be referred to by the figures they appear in.

[d] Although the necessary groups are available, no statistical test of the R*I interaction is reported.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	autobio-graphical narratives	positive-valence event	number of 'details' (cued recall)	stress	S-A	N	^	Y	Y	N	N	N	N	U[d]	?	?	?
Forcato et al. (2010)	paired-associates	:	L1 errors (cued-recall)	new learning	S-A	C	12	Y	Y	?	Y	?	?	?	?	?	?
Strange et al. (2010)	list learning	E3	correct (cued recall)	negative-valence stimuli	S-A	D	16	Y	Y	?	?	Y	?	?	?	?	Y
	^	^	^	neutral-valence stimuli	^	N	^	Y	Y	?	?	N	?	?	:	:	:
	list learning	E4	correct (cued recall)	negative-valence stimuli	S-A	D	11	Y	Y	?	?	Y	?	?	?	Y	?
	^	^	^	neutral-valence stimuli	^	N	^	Y	Y	?	?	N	?	?	:	:	:
	list learning	E5	correct (cued recall)	negative-valence stimuli	S-A	D	14	Y	Y	?	?	Y	?	?	?	?	Y
	^	^	^	neutral-valence stimuli	^	N	^	Y	Y	?	?	N	?	?	:	:	:
Schiller et al. (2010)	fear-conditioning	E1	skin conductance response	extinction	S-A	D	20	Y	Y	?	U[e]	?	?	U[e]	U[e]	?	U[e]
	fear-conditioning	E2	skin conductance response	extinction	S-A	D	18	Y	Y	?	U[e]	?	?	U[e]	?	?	?

[e] Although the necessary groups are available, no statistical test of the R*I interaction is reported (see subsection 2.2.3).

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Marin et al. (2010)	story learning	negative-valence story	correct (free recall)	stress	S-A	S	16	Y	Y	?	Y	N[f]	?	?	?	N	?
	story learning	neutral-valence story	correct (free recall)	stress	S-A	N	^	Y	Y	?	N	N	?	?	:	:	:
Soeter & Kindt (2011)	fear-conditioning	E2	eyeblink startle reflex	extinction	S-A	N	?[g]	Y	Y	?	Y	?	?	?	:	:	N
	^	^	skin conductance response	^	^	N	^	Y	Y	?	N	?	?	?	:	:	:
	^	^	US expectancy	^	^	N	^	Y	Y	?	N	?	?	?	:	:	:
	^	^	subjective anxiety	^	^	N	^	Y	Y	?	N	?	?	?	:	:	:
Chan et al. (2011)	post-event misinformation	:	recall correct (cued recall)	new learning	S-A	N	39	N	Y	?	U[h]	U[h]	U[h]	U[h]	N	:	:
	^	:	recall misinformation (cued recall)	^	^	N	^	N	Y	?	Y	Y	Y	Y	N	:	:

[f]Significant group difference on immediate test but not delayed test.

[g]Per group sample sizes do not appear to be reported.

[h]A similar pattern was observed as in Chan et al. (2009; see note accompanying that study). The pattern was also apparent in a delayed intervention group ($R^{+}I^{D}$) and so cannot be attributed to disruption of reconsolidation.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Hupbach et al. (2011)	list learning	:	correct (free recall)	new learning	G-P	N	11	Y	Y	?	N	?	?	?	:	:	:
	^	:	intrusions (free recall)	new learning	G-P	C	^	Y	Y	?	Y	?	?	?	?	Y	?
Diekelmann et al. (2011)	object location	awake during reminder	correctly recalled locations	new learning	G-P	D	12	N	Y	?	Y	?	?	?	?	N[i]	?
	object location	asleep during reminder	correctly recalled locations	new learning	G-P	N	12	N	Y	?	N	?	?	?	?	N[i]	?
Coccoz et al. (2011)	paired-associates	E1	correct (cued-recall)	stress	S-A	S	11	Y	Y	?	?	Y	?	?	?	Y	?
	paired-associates	E2 overt response (retrieval)	correct (cued-recall)	stress	S-A	S	11	Y	Y	?	Y	Y	?	?	?	?	?
	paired-associates	E2 no overt response (retrieval)	correct (cued-recall)	stress	S-A	N	11	Y	Y	?	N	Y	?	?	:	:	:
Forcato et al. (2011)	paired-associates	E1-A	L1 errors (cued-recall)	none	S-A	S	13	Y	Y	?	?	?	?	?	?	?	?
	paired-associates	E1-B	L1 errors (cued-recall)	none	S-A	N	13	Y	Y	?	N	?	?	?	?	?	?

[i] participants were only tested 30m after the intervention (reconsolidation effects are not expected to emerge until after a delay, e.g., 24h)

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Wichert et al. (2011)	list learning	1d training-reminder interval	number recalled (free recall)	new learning	S-A	N	12	Y	Y	?	N	Y	Y	N	:	:	:
	list learning	7d training-reminder interval	number recalled (free recall)	new learning	S-A	D	12	Y	Y	?	N[j]	Y	Y	N	?	?	?
	list learning	28d training-reminder interval	number recalled (free recall)	new learning	S-A	N	12	Y	Y	?	N	Y	Y	N	:	:	:
Finn et al. (2011)	paired-associates	E1: reminder-intervention delay (0s)	correct (cued recall)	negative-valence stimuli	S-A	S	40	N	Y	?	?	Y	?	?	?	?	?
	paired-associates	E2: reminder-intervention delay (2s)	correct (cued recall)	negative-valence stimuli	S-A	S	56	N	Y	?	?	Y	?	?	?	?	?
	paired-associates	E3: reminder is restudy	correct (cued recall)	negative-valence stimuli	S-P	N	61	N	Y	?	?	N	?	?	:	:	:
Finn et al. (2012)	paired-associates	E1: post-retrieval intervention	correct (cued recall)	negative-valence stimuli	S-A	S	53	N	Y	?	?	Y	?	?	?	?	?
	paired-associates	E1: post-retrieval intervention	correct (cued recall)	positive-valence stimuli	S-A	N	60	N	Y	?	?	N	?	?	:	:	:

[j] The authors argue in favour of reconsolidation theory by suggesting that there is a meaningful difference between the R+I+ and R-I+ groups. However the difference was not statistically significant ($p = .08$).

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	paired-associates	E1: pre-retrieval intervention	correct (cued recall)	negative-valence stimuli	S-A	N	57	N	N	?	?	N	?	?	:	:	:
	paired-associates	E2: successful retrieval-practice	correct (cued recall)	negative-valence stimuli	S-A	S	104	N	Y	?	?	Y	?	?	?	?	?
	paired-associates	E2: unsuccessful retrieval-practice	proportion correct (cued recall)	negative-valence stimuli	S-A	S	107	N	Y	?	?	Y	?	?	?	?	?
Potts & Shanks (2012)	paired-associates	:	correct (cued recall)	new learning	S-A	N	18	Y	Y	Y	N	Y	Y	N[k]	:	:	:
Chan et al. (2012)	post-event misinformation	cued-recall test	correct (cued recall)	new learning	S-A	N	40	N	Y	?	N	?[l]	?[l]	?[l]	:	:	:
	^	source attribution test	source misattributions (control items)	^	^	N	^	N	Y	?	Y	?[l]	?[l]	?[l]	:	:	:
	^	source attribution test	source misattributions (misled items)	^	^	C	^	N	Y	?	Y	?[l]	?[l]	?[l]	?	N	?

[k] there was a significant R*I interaction but follow-up tests indicated that the effects were not in the direction expected by reconsolidation theory e.g., the R-I+ group performed substantially worse than the R+I+ group

[l] Although $R^{+}I^{-}$ and $R^{-}I^{-}$ groups were included in the design, the authors rightly point out that comparisons with these groups were not appropriate due to a counter-balancing issue that had a substantial influence on baseline (i.e., pre-reminder) performance.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	^	non-overlapping source attribution test	source misattributions (control items)	^	^	N	^	N	Y	?	N	?[l]	?[l]	?[l]	:	:	:
	^	non-overlapping source attribution test	source misattributions (misled items)	^	^	N	^	N	Y	?	N[m]	?[m]	?[m]	?[m]	:	:	:
Golkar et al. (2012)	fear-conditioning	E1: fear-relevant stimuli	eyeblink startle reflex	extinction	S-A	N	19	Y	Y	?	N	?	?	?	:	:	N
	^	^	skin conductance response	^	^	N	^	Y	Y	?	N	?	?	?	:	:	N
	fear-conditioning	E2: fear-irrelevant stimuli	skin conductance response	extinction	S-A	N	20	Y	Y	?	N	?	?	?	:	:	N

[m] See above note. Additionally there was an effect of retrieval but in the opposite direction expected by reconsolidation theory.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Agren et al. (2012a) and Bjorkstrand et al. (2015)[n]	fear-conditioning	:	skin conductance response	extinction	S-A	D	11	Y	Y	?	?	?	?	?	N[o]	?	N[o]
Agren et al. (2012b)	fear-conditioning	:	skin conductance response	extinction	S-A	D	32	Y	Y	?	?	?	?	?	Y	?	?
Oyarzun et al. (2012)	fear-conditioning	:	skin conductance response	extinction	S-A	D	17	Y	Y	?	Y	?	?	?	?	?	?
Xue et al. (2012)	drug-related addiction	:	subjective heroin craving	extinction	S-A	D	22	Y	Y	?	Y	?	?	?	Y	?	Y
	^	:	blood pressure	^	^	N	^	Y	Y	?	N	?	?	?	N	?	N
	^	:	heart rate	^	^	N	^	Y	Y	?	N	?	?	?	N	?	N
Kindt & Soeter (2013)	fear-conditioning	:	eyeblink startle reflex	extinction	S-A	N	?[p]	Y	Y	?	N	?	?	?	:	:	N

[n]Note that Bjorkstrand et al. is an 18-month follow up to Agren et al. and so they are considered together here.

[o]In both experiments, a significant effect is observed in one group (6h) but not the other (10m), but there is no statistical test of the difference between the groups. A re-analysis of the data reveals no significant effect of group (see subsection 2.2.3).

[p]Per group sample sizes do not appear to be reported.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	^	:	skin conductance response	^	^	N	^	Y	Y	?	N	?	?	?	:	:	N
	^	:	US expectancy	^	^	N	^	Y	Y	?	N	?	?	?	:	:	N
Warren et al. (2013)	fear-conditioning	US-expectancy recorded	eyeblink startle reflex	extinction	S-A	?[q]	12	Y	Y	?	N	?	?	?	:	:	Y
	^	^	US expectancy	^	^	?[q]	^	Y	Y	?	N	?	?	?	:	:	Y
	fear-conditioning	US-expectancy not recorded	eyeblink startle reflex	extinction	S-A	N	20	Y	Y	?	N	?	?	?	:	:	N
Schiller et al. (2013)	fear-conditioning		skin conductance response	extinction	S-A	D	19	Y	Y	?	Y	?	?	?	?	?	?
St Jacques & Schacter (2013)	autobio-graphical museum tour	E1	hits (yes/no recognition)	new learning	S-P	S	42	Y	Y	?	U[r]	?	?	?	?	?	?

[q]The authors note some effects of the reminder manipulation on reinstatement, but not on spontaneous recovery. It is not clear if they find this compelling evidence for a reconsolidation effect or not.

[r]the statistical approach used here may be problematic.There were three types of within-subjects reminder condition (Reminder Match; Reminder Mismatch; and No Reminder) where ‘match’ referred to whether events presented in the reminder followed the same or different temporal order of the actual Day 1 event. For analysis, the scores for the No Reminder condition are subtracted from the R-Match and R-Mismatch conditions, then these two conditions are compared to each other (via ANOVA). Therefore, it is not entirely clear if either R+ condition differed significantly from the R- condition, only that they differed from each other. Nevertheless, the numerical pattern of the data is consistent with a Reminder-contingent effect.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	^	^	false alarms (yes/no recognition)	^	^	C	^	Y	Y	?	U[r]	?	?	?	?	?	?
	autobio-graphical museum tour	E2	hits (yes/no recognition)	new learning	S-P	S	41	Y	Y	?	U[r]	?	?	?	?	?	?
	^	^	false alarms (yes/no recognition)	^	^	C	^	Y	Y	?	U[r]	?	?	?	?	?	?
St Jacques et al. (2013)	autobio-graphical museum tour	:	hits (yes/no recognition)	new learning	S-P	S	35	Y	Y	?	Y	?	?	?	?	?	?
	^	:	false alarms (yes/no recognition)	^	^	C	^	Y	Y	?	Y	?	?	?	?	?	?
Chan & LaPaglia (2013)	post-event misinforma-tion	E1	hits - false alarms (true/false recognition)	new learning	S-A	D	70	N	Y	N	Y	Y	Y	U[s]	U[s]	N	?
	post-event misinforma-tion	E3	hits - false alarms (true/false recognition)	new learning	S-A	D	32	Y	Y	N	Y	Y	Y	?[t]	?	N	?

[s]Although the necessary groups are available, no statistical test of the R*I interaction is reported. Additionally, although a $R^{+}I^{D}$ was included, no statistical comparison was made with the non-delayed intervention groups.

[t]Although the necessary groups are available, no statistical test of the R*I interaction is reported.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	post-event misinformation	E4	hits (source-free old/new recognition)	new learning	S-A	D	36	N	Y	N	Y	Y	Y	?[t]	?	N	?
	post-event misinformation	E6	hits - false alarms (true/false recognition)	new learning	S-A	D	33	Y	Y	N	Y	Y	Y	?[t]	?	?	?
Pashler et al. (2013)	semantic fact updating	E2	correct (cued recall)	new learning	S-A	N	69	Y	Y	?	N[u]	?	?	?	:	:	:
	^	^	correct (cued recall)	^	S-P	N	^	Y	Y	?	N[u]	?	?	?	:	:	:
Coccoz et al. (2013)	paired-associates	E1:Training-reminder delay, 7d	correct (cued recall)	stress	S-A	S	8	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E1: Training-reminder delay, 21d	correct (cued recall)	stress	S-A	N	6	Y	Y	?	?	N	?	?	?	?	?
Forcato et al. (2013)	paired-associates	E1: Training-Reminder delay, 1d, one reminder	errors (cued-recall)	none	S-A	S	12	Y	Y	?	?	?	Y	?	?	?	?

[u] In fact, the researchers found that retrieval practice and restudy (followed by new learning) enhanced rather than disrupted subsequent performance of Day 1 facts.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	paired-associates	E1: Training-Reminder delay, 1d, two-reminders	errors (cued-recall)	none	S-A	S	12	Y	Y	?	?	?	Y	?	?	?	?
	paired-associates	E2: One reminder	errors (cued-recall)	new learning	S-A	D	13	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E2: Two-reminders	errors (cued-recall)	new learning	S-A	D	13	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E3: Training-Reminder delay, 8d, one reminder	errors (cued-recall)	none	S-A	N	10	Y	Y	?	N	?	?	?	?	?	?
	paired-associates	E3: Training-Reminder delay, 8d, two reminders	errors (cued-recall)	none	S-A	N	10	Y	Y	?	N	?	?	?	?	?	?
	paired-associates	E3: Training-Reminder delay, 8d, four reminders	errors (cued-recall)	none	S-A	N	10	Y	Y	?	N	?	?	?	?	?	?
	paired-associates	E4: Training-Reminder delay, 8d, one reminder	errors (cued-recall)	new learning	S-A	N	10	Y	Y	?	?	?	N	?	?	?	?
Gershman et al. (2013)	list learning	:	source misattributions	new learning	G-A	C	14	Y	Y	?	?	?	?	?	?	?	?

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Wichert et al. (2013a)	list learning	zero pre-Day 2 retrievals	hits (old/new recognition)	new learning	S-A	D	14	Y	Y	?	Y	Y	Y	N[v]	?	?	?
	^	zero pre-Day 2 retrievals	false alarms (old/new recognition)	^	^	N	^	Y	Y	?	N	N	Y	N	:	:	:
	list learning	one pre-Day 2 retrieval	hits (old/new recognition)	new learning	S-A	D	14	Y	Y	?	N	Y	Y	N[w]	?	?	?
	^	one pre-Day 2 retrieval	false alarms (old/new recognition)	^	^	N	^	Y	Y	?	N	N	Y	N	:	:	:
	list learning	three pre-Day 2 retrievals	hits (old/new recognition)	new learning	S-A	D	13	Y	Y	?	N	Y	Y	N[w]	?	?	?
	^	three pre-Day 2 retrievals	false alarms (old/new recognition)	^	^	N	^	Y	Y	?	N	N	Y	N	:	:	:
Wichert et al. (2013b)	list learning	stronger intervention	false alarms (old/new recognition)	new learning	S-A	C	12	Y	Y	?	Y[x]	Y	Y	Y	?	?	?

[v] In the main analysis the reminder-intervention interaction is significant, leading the authors to conclude that there was a reconsolidation effect. However, in an analysis that only included items successfully recalled during the Day 2 Reminder Stage (and thus presumably those items that theoretically started to undergo reconsolidation), the reminder-intervention interaction was no longer significant.

[w] The authors tentatively suggest that there is still evidence for a reconsolidation effect based on numerically lower performance in the R^+I^+ group relative to other experimental groups. However, the critical R*I interaction was not statistically significant.

[x] However, note that the difference between the R^+I^+ strong encoding condition and the R-I+ weak encoding condition was marginally significant (p =.08)

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	list learning	weaker intervention	false alarms (old/new recognition)	new learning	S-A	N	12	Y	Y	?	N	Y	Y	N	:	:	:
	^	strong and weak interventions	hits (old/new recognition)	^	^	N	^	Y	Y	?	N	N	Y	N	:	:	:
Beukelaar et al. (2014)	motor sequence learning	"long" reactivation (3x30s)	correct / inter-tap interval	new learning	S-A	N	12	Y	Y	N	?	N	?	?	?	?	?
	motor sequence learning	"short" reactivation (∼7.5s)	correct / inter-tap interval	new learning	S-A	D	12	Y	Y	Y	?	Y	?	?	?	?	?
	motor sequence learning	"medium" reactivation (1x60s)	correct / inter-tap interval	new learning	S-A	N	12	Y	Y	N	?	?	?	?	?	?	?
	motor sequence learning	"medium" reactivation (1x30s)	correct / inter-tap interval	new learning	S-A	D	12	Y	Y	Y	?	?	?	?	?	?	?
	motor sequence learning	"medium" reactivation (∼15.6s)	correct / inter-tap interval	new learning	S-A	D	12	Y	Y	Y	?	?	?	?	?	?	?
Pine et al. (2014)	instrumental conditioning	:	correct (instrumental responses)	new learning	S-A	D	19	Y	Y	?	Y	?	?	?	?	?	?
Stein-furth et al. (2014)	fear-conditioning	Training - Reminder delay 1d	skin conductance response	extinction	S-A	D	20	Y	Y	?	Y	?	?	?	?	?	?

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	fear-conditioning	Training - Reminder delay 7d	skin conductance response	extinction	S-A	D	20	Y	Y	?	Y	?	?	?	?	?	?
Drexler et al. (2014)	fear-conditioning	:	skin conductance response	extinction	S-A	N	?[y]	Y	Y	?	N	?	?	?	:	:	N
	^	:	US expectancy	^	^	N	^	Y	Y	?	N	?	?	?	:	:	N
Johnson & Casey (2015)	fear-conditioning	adults	skin conductance response	extinction	S-A	D	18	Y	Y	?	Y	?	?	?	?	?	?
	fear-conditioning	adolescents	skin conductance response	extinction	S-A	D	19	Y	Y	?	Y	?	?	?	?	?	?
Jones et al. (2015)	list learning	young adults	correct (free recall)	new learning	G-P	N	13	Y	Y	?	N	?	?	?	:	:	:
	^	^	intrusions (free recall)	^	^	C	^	Y	Y	?	Y	?	?	?	?	?	?
	list learning	older adults	correct (free recall)	new learning	G-P	N	20	Y	Y	?	N	?	?	?	:	:	:
	^	^	intrusions (free recall)	^	^	N	^	Y	Y	?	N	?	?	?	:	:	:
Wirkner et al. (2015)	list learning	:	hits - false alarms (old/new recognition)	new learning	S-P	D	20	Y	Y	?	N	Y	Y	?	?	?	?

[y] Per group sample sizes do not appear to be reported.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	⌢	:	event-related potentials	⌢	⌢	D	⌢	Y	Y	?	N	Y	Y	?	?	?	?
Yacoby et al. (2015)	video watching	:	correct (multiple-choice recognition)	new learning	S-A	D	34	Y	Y	N	N[z]	?	?	?	N[z]	?	?
Hupbach (2015)	list learning	retrieval practice	intrusions (free recall)	new learning	S-A	N	13	Y	Y	?	N	?	?	?	:	:	:
	list learning	subtle reminder	intrusions (free recall)	new learning	G-A	C	12	Y	Y	?	Y	?	?	?	?	?	?
	⌢	retrieval practice and subtle reminder	correct (free recall)	⌢	⌢	N	⌢	Y	Y	?	N	?	?	?	:	:	:
Kredlow & Otto (2015)	traumatic autobio-graphical memory	neutral-valence story	details' (free recall)	new learning	S-A	N	25	Y	Y	?	?	N	?	?	?	?	?
	traumatic autobio-graphical memory	negative-valence story	details' (free recall)	new learning	S-A	D	23	Y	Y	?	?	Y	?	?	?	?	?

[z] Although no significant difference between the R^+I^+ group and the other experimental groups was observed, the researchers identified a significant correlation in the R^+I^+ group between subjective 'feeling of knowing' (FOK) ratings during the Reminder Stage and the degree of performance impairment between baseline and final test. On the basis that FOK may be a proxy for the extent of trace reactivation, the researchers suggested that higher levels of FOK/reactivation might be necessary for post-retrieval new learning to effectively disrupt reconsolidation.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	traumatic autobio-graphical memory	positive-valence story	details' (free recall)	new learning	S-A	N	24	Y	Y	?	?	N	?	?	?	?	?
St Jacques et al. (2015)	autobio-graphical museum tour	:	hits (yes/no recognition)	new learning	S-P	S	32	Y	Y	?	Y	?	?	?	?	?	?
	^	:	false alarms (yes/no recognition)	^	^	C	^	Y	Y	?	Y	?	?	?	?	?	?
James et al. (2015)	traumatic autobio-graphical memory	E1	intrusive-memory frequency	behavioural task	S-P	D	26	Y	Y	?	?	?	Y	?	?	?	?
	^	^	hits (true/false recognition)	^	^	N	^	Y	Y	?	?	?	N	?	:	:	:
	traumatic autobio-graphical memory	E2	intrusive-memory frequency	behavioural task	S-P	D	18	Y	Y	?	Y	Y	Y	N[a]	?	?	?
	^	^	hits (true/false recognition)	^	^	N	^	Y	Y	?	N	N	N	?	:	:	:

[a] Although the necessary groups are available, no statistical test of the R*I interaction is reported. However, re-analysis of the data indicates that the interaction is not statistically significant (see subsection 2.2.2).

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Hardwicke et al. (2016, Chapter 3)	motor sequence learning	E1-E4	accuracy	new learning	S-A	N	48	Y	Y	N	?	?	?	?	:	:	:
	^	^	speed	^	^	N	^	Y	Y	N	?	?	?	?	:	:	:
	declarative sequence learning	E5-E7	sequence similarity	new learning	S-A	N	32	Y	Y	Y	N	?	?	?	:	:	:
Fernandez et al. (2016)	paired-associates	E1: with prior social stress	L1 errors (cued-recall)	new learning	S-A	D	12	Y	Y	?	N	Y	?	?	?	?	?
	paired-associates	E1: without prior social stress	L1 errors (cued-recall)	new learning	S-A	N	12	Y	Y	?	?	Y	?	?	?	?	?
	paired-associates	E2: with prior social stress	L1 errors (cued-recall)	new learning	S-A	D	12	Y	Y	?	N	Y	?	?	?	?	?
	paired-associates	E3: one reminder	L1 errors (cued-recall)	new learning	S-A	S	12	Y	Y	?	?	?	Y	?	?	?	?
	paired-associates	E3: two reminders	L1 errors (cued-recall)	new learning	S-A	S	12	Y	Y	?	?	?	Y	?	?	?	?
Klucken et al. (2016)	fear-conditioning	:	skin conductance response	extinction	S-A	N	70	Y	Y	?	N	?	?	?	:	:	N
Fricchione et al. (2016)	fear-conditioning	:	skin conductance response	extinction	S-A	N	21	Y	Y	?	N	?	?	?	:	:	N

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Hardwicke et al. (Chapter 4)	post-event misinformation	E8: 'original' test	correct (recognition)	new learning	S-A	N[b]	35	Y	Y	?	?	Y	?	?	:	:	:
	post-event misinformation	E8: 'modified' test	correct (recognition)	new learning	S-A	N	24	Y	Y	?	?	N	?	?	:	:	:
	post-event misinformation	E9: 'true/false' test	hits minus false alarms (recognition)	new learning	S-A	N[b]	65	Y	Y	?	?	Y	?	?	:	:	:
	^	E9: 'modified' test	correct (recognition)	^	^	N	^	Y	Y	?	?	N	?	?	:	:	:
	post-event misinformation	E10: 'modified' test, retrieval practice	correct (recognition)	new learning	S-A	N	44	Y	Y	?	N	N	?	?	:	:	:
	post-event misinformation	E10: 'modified' test, subtle reminder	correct (recognition)	new learning	G-A	N	44	Y	Y	?	N	N	?	?	:	:	:

[b] Although an intervention-dependent impairment was observed, the effect was eliminated when a more sensitive test was used, see row below and Chapter 4 for details.

Table 2.5

Study	Paradigm	Subgroup	DV	Intervention	R	Effect	N	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Hardwicke et al. (Chapter 5)	semantic fact updating	E11: Restudy	correct (cued recall)	new learning	S-P	N	59	Y	Y	?	N[c]	?	?	?	:	:	:
	^	E11: Retrieval-practice	correct (cued recall)	^	S-A	N	^	Y	Y	?	N[c]	?	?	?	:	:	:

[c] There was an effect of reminder condition, but in the opposite direction to the predictions of reconsolidation theory.

Chapter 3

Investigating reconsolidation in a sequence learning paradigm: Attempted replications of Walker et al. (2003)

A particularly prominent finding reported by Walker et al. (2003, also referred to here as the "original study") is widely cited as a convincing demonstration of reconsolidation-mediated memory updating in humans (e.g., Tronson & Taylor, 2007; J. L. Lee, 2009; Nader & Hardt, 2009; Schiller & Phelps, 2011; Schwabe et al., 2014). The results are especially compelling because the experiment conformed to the canonical 3-day reconsolidation protocol (Figure 3.1) typically used in non-human animal studies, thus meeting several key criteria necessary for a robust investigation of reconsolidation (see Chapter 2, Nader & Hardt, 2009; Schwabe et al., 2014; Tronson & Taylor, 2007).

On day 1, participants used a computer keyboard to repeatedly tap a simple sequence of on-screen digits (e.g., 41342). Speed and accuracy improvements were observed as participants learned this initial ('Old') sequence. On day 2, participants in the Reminder group (n = 16) practiced the Old Sequence immediately before learning a New Sequence. The No-Reminder group did not practice the Old Sequence before new learning. The No-Intervention group practiced the Old Sequence but did not learn a New Sequence. On day 3, sequence performance was tested for all groups. The key finding was that the Reminder group's Old Sequence accuracy suffered a substantial decline ($\sim 57\%$) between the Reminder Stage and the Test Stage, although only minor decrements were observed on the speed measure ($\sim 2\%$). By contrast, improvements in accuracy and speed between Training and Test Stages were observed in the No-Reminder and No-Intervention groups. Therefore, it would appear that the accuracy impairment in the Reminder group was contingent on both the reminder and intervention as demonstrated in similar non-human animal studies (Nader et al., 2000a) and widely accepted as evidence for reconsolidation (Nader & Hardt, 2009; Schwabe et al., 2014; Tronson & Taylor, 2007). Consistent with the view that the Old Sequence memory trace had been rewritten by the new learning (J. C. K. Chan & LaPaglia, 2013; Schiller et al., 2010), the authors suggested that reconsolidation may have "functional significance," allowing the "continued refinement and reshaping of previously learned movement skills" (Walker et al., 2003, p. 618).

However, from the perspective of the aforementioned storage–retrieval debate (Chapter 1), this interpretation should be viewed with caution, especially as retrieval-deficit explanations were not explored. For example, it was not clear whether the effect endured beyond the 3-day study period, or showed propensity for recovery under favorable retrieval conditions (see Chapter 2, criterion 10, Bouton, 2002), effects that have been observed in several investigations of reconsolidation with non-human animals (e.g., Eisenberg & Dudai, 2004; Lattal & Abel, 2004; Power, Berlau, McGaugh, & Steward, 2006). In the present study, we initially sought to replicate and extend the reported reconsolidation effect (Walker et al., 2003, Group 7) by examining whether it could be accounted for by retrieval-deficits rather than the storage-deficit mechanisms outlined under reconsolidation theory (our investigation does not address other findings, unrelated to reconsolidation, reported in the same article). We conducted a replication battery (Rosenthal, 1990) consisting of both 'direct replications' (Simons, 2014) that followed the methodology of the original study as closely as possible, and 'conceptual replications' (Schmidt, 2009) that manipulated key task parameters to explore the broader validity of the reconsolidation-updating theory.

3.1 General Methods

All experimental software and materials, data, and analysis code pertaining to this chapter have been made publicly available on the Open Science Framework (https://osf.io/ry8fw). All data exclusions, manipulations, and measures conducted during this study are reported.

Participants were recruited from the University College London (UCL) mixed-occupation subject pool and received either monetary compensation or course credits. All participants reported that they were right-handed and had no history of neurological, psychiatric, or sleep disorder. All participants provided informed consent and the study was approved by the local UCL ethics committee.

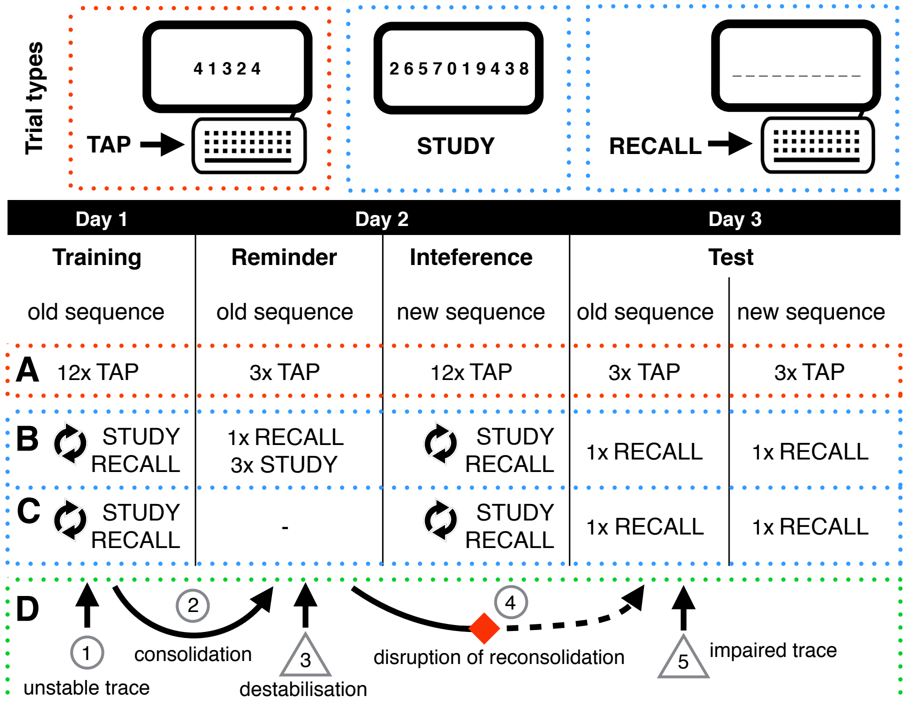


Figure 3.1: Reconsolidation sequence learning paradigm for Walker et al. (2003) and direct replications (**A**, red boundary, Experiments 1-4), conceptual replications (blue boundary) with reminder condition (**B**, Experiments 5-6) and without reminder condition (**C**, Experiment 7), and hypothesized underlying mechanisms and events predicted by reconsolidation theory (**D**, green boundary). Critical time points for calculation of the reconsolidation score (RS) are indicated by triangle symbols. See main text for details.

3.2 Experiments 1-4

We initially intended to run a single direct replication in order to establish the reliability of the effect under scrutiny before conducting additional manipulations. However, to foreshadow our findings, the complete absence of a reconsolidation effect precluded any further investigation of a retrieval-deficit account. Instead, we made several attempts to reproduce the target effect in repeated direct replications ($N = 64$) using our own software (Experiment 1), software provided by the original researchers (Experiment 2), and under conditions intended to increase task difficulty (Experiments 3 and 4).

3.2.1 Methods

3.2.1.1 Participants

Sixteen participants took part in each of the four direct replication experiments, affording a total sample size of 64 individuals (49 females; age $Mdn = 22$ years, range = 18–54 years). Two additional participants were excluded for typing an incorrect sequence at the Reminder Stage, and four additional participants did not complete all three stages of the study.

3.2.1.2 Design

Participants performed a ‘finger-tapping’ sequence learning task in three discrete sessions taking place on consecutive days (Figure 3.1). Two five-digit sequences (X: 4–1–3–2–4; Y: 2–3–1–4–2) were assigned to be the Old Sequence and the New Sequence in counterbalanced order. On day 1, participants completed 12 Old Sequence trials (Training). On day 2, participants performed three Old Sequence trials (Reminder) immediately before 12 New Sequence trials (Interference). On day 3, participants completed three trials of both the Old Sequence and the New Sequence in counterbalanced order (Test). The dependent variables (for details see Appendix A) were the number of sequences completed during each 30s trial (‘speed’) and the ratio of errors to speed (‘accuracy’: $1 - \frac{\text{errors}}{\text{speed}}$).

Sequence order (X or Y; i.e., whether the Old Sequence was designated as 4–1–3–2–4 or 2–3–1–4–2, with the remaining sequence being assigned as the New Sequence) and test order (A or B; i.e., whether the Old or New Sequence came first on the Day 3 Test) were counterbalanced ($n = 8$ per condition), except in Experiment 2 where researcher error led to unbalanced conditions (test order: A = 12, B = 4; sequence order: X = 7, Y = 9). Neither test order nor sequence order influenced the presence or absence of reconsolidation effects (see Appendix B).

3.2.1.3 General Procedure

Unless otherwise stated below (see subsubsection 3.2.1.4), the following procedures were used in all direct replications and precisely matched those reported in the original study. Ambiguous or missing information was clarified through contact with the senior author of the original research team. Participants were seated in front of a computer screen in a quiet room and used the four fingers of their left (non-dominant) hand to respond using the four top-row numeric keys 1, 2, 3, and 4 of a standard keyboard. The task involved repeatedly tapping a five-element sequence that was displayed on the screen for 30s (including on 'test' trials), followed by 30s of rest during which the sequence was absent. Key presses were acknowledged with white dots that accumulated on screen, but there was no feedback regarding response accuracy. A 30s countdown timer was displayed during the rest phase to signal the approaching test phase. During the tapping phase, the screen background was green, and during the rest phase it was red. Participants were instructed to 'tap out the sequence as quickly and accurately as possible.' There was no within- or between-subjects timing variability in the original study because all sessions were conducted at 1:00 PM. In the present experiments, there was also no within-subject variability: participants completed sessions at precise 24 hour intervals (±15 min); however, session times varied between participants (9:00 AM to 6:00 PM).

3.2.1.4 Procedural variations

Each experiment had minor variations from the general procedure outlined above. Unlike the other experiments, in Experiment 1 the sequence remained on screen during rest trials, there was no countdown timer, and the background color was invariant throughout. Key presses were acknowledged with the transient display of white dots arranged in a row that corresponded to the horizontal order of the physical keys. Experiments 1, 3, and 4 were executed in Python code (https://osf.io/g9n87/) developed by T.E.H., whereas Experiment 2 was run from an executable file provided by the original research team.

In the original study and Experiments 1 and 2, participants were instructed to tap the sequence, '…as quickly and accurately as possible.' In Experiments 3 and 4, this instruction was modified to read '…as quickly as you can. Try not to make errors, but overall you should emphasize speed over accuracy.' The phrase 'tap as quickly as you can!' was also displayed continuously on screen during test phases in Experiments 3 and 4. In Experiment 4, the keyboard was positioned in an adapted box file such that the participant was unable to view their hand during task performance. Tactile markers were placed on the response keys to prevent the participants' hand shifting to the incorrect keys. Participants were allowed to lift the lid of the box file during rest phases so they could stretch their fingers and ensure the hand was correctly positioned before closing the lid and starting the next trial.

3.2.2 Results

Statistical significance was defined at the .05 level. Minor procedural differences between the replications and the original study (variability in participant age and time of testing) were ruled out as potential confounds through additional analyses (Appendix B).

3.2.2.1 Training Stage (Old Sequence)

To establish whether there were performance gains during the Old Sequence Training Stage, we used two separate 4 × 12 mixed-factorial ANOVAs for accuracy and speed with one between-subjects variable (Experiment: 1–4) and one within-subjects variable (Trial: 1–12). *Accuracy* (Figure 3.2A, Training) increased numerically between trial 1 ($M = 0.806$, $SD = 0.238$) and trial 12 ($M = 0.879$, $SD = 0.132$), although neither the linear [$F(1,60) = 2.96$, $p = 0.091$] nor quadratic [$F(1,60) = 0.105$, $p = 0.747$] trend was significant. There was no significant interaction between experiment and either linear [$F(3,60) = 0.580$, $p = 0.630$] or quadratic [$F(3,60) = 1.203$, $p = 0.316$] trend of trial. *Speed* (Figure 3.2B, Training) increased numerically between trial 1 ($M = 14.200$, $SD = 6.40$) and trial 12 ($M = 21.238$, $SD = 6.356$). Linear [$F(1,60) = 95.398$, $p < 0.001$] and quadratic [$F(1,60) = 12.908$, $p = 0.001$] trends were both statistically significant. There was no significant interaction between experiment and linear trend of trial, $F(3,60) = 1.371$, $p = 0.260$, but the interaction between experiment and quadratic trend of trial was significant, $F(3,60) = 3.209$, $p = 0.029$.

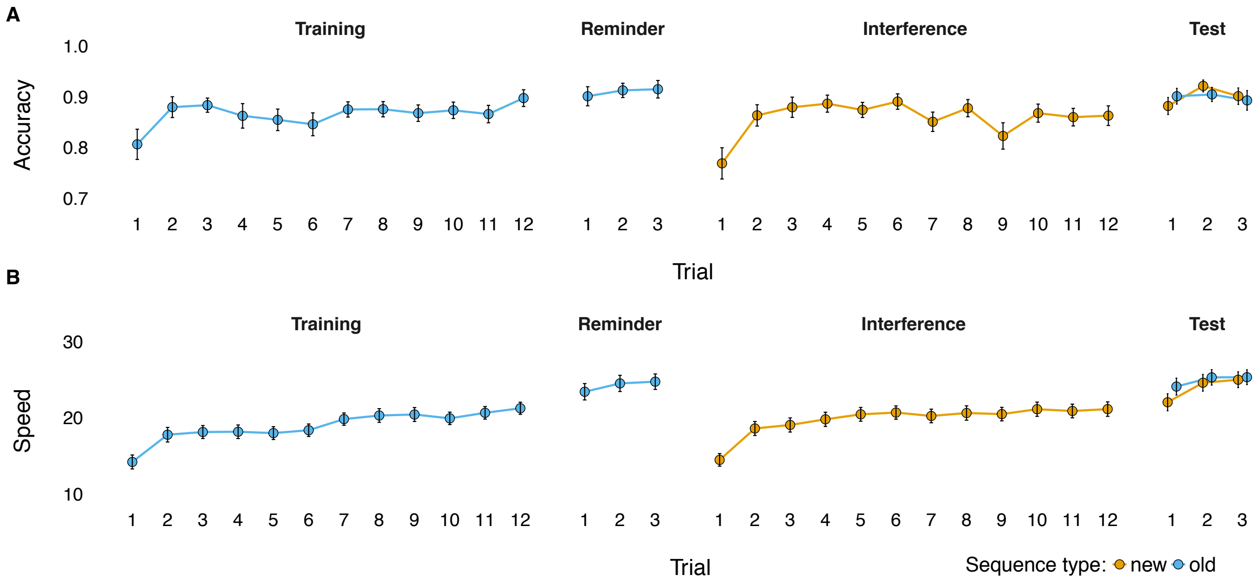


Figure 3.2: Experiments 1-4 (pooled) timeline showing mean accuracy (**A**; number of errors made relative to the number of complete sequences achieved) and mean speed (**B**; number of complete sequences achieved) by stage (Training, Reminder, Interference, and Test), trial, and sequence type. Error bars show ±SEM.

3.2.2.2 Interference Stage (New Sequence)

The same ANOVA design was used to assess changes in New Sequence performance across the Interference Stage. *Accuracy* (Figure 3.2A, Interference) increased numerically between trial 1 (M = 0.769, SD = 0.246) and trial 12 (M = 0.862, SD = 0.155). Across trials, there was a significant quadratic trend, $F(1,60) = 7.651$, $p = 0.008$. The linear trend was not significant, $F(1,60) = 1.087$, $p = 0.301$. There was no significant interaction between experiment and quadratic trend of trial, $F(3,60) = 1.254$, $p = 0.298$, or linear trend of trial, $F(3,60) = 1.327$, $p = 0.274$. *Speed* (Figure 3.2B, Interference) increased numerically between trial 1 (M = 14.466, SD = 6.593) and trial 12 (M = 21.128, SD = 7.588). A linear trend, $F(1,60) = 83.075$, $p < 0.001$, and quadratic trend, $F(1,60) = 58.072$, $p < 0.001$, were both significant. There was no interaction between experiment and linear trend, $F(3,60) = 1.137$, $p = 0.342$, or quadratic trend $F(3,60) = 1.362$, $p = 0.263$.

3.2.2.3 Between-Stage Comparisons

In the original study, two comparisons were made to examine overnight changes in sequence performance. *Overnight Score Old* (OSO) was the percentage change between Old Sequence Training (trials 10–12 only[22]) and Old Sequence Reminder (all three trials). *Overnight Score New* (OSN) was the percentage change between New Sequence Interference (trials 10–12 only) and New Sequence Test (all three trials). To establish whether the overnight scores varied between experiments, we used a series of one-way ANOVAs with Experiment (1–4) as a between-subjects factor and overnight score (separately for old/new and separately for accuracy/speed) as a dependent variable. For *accuracy*, there was no significant main effect of Experiment for OSO, $F(3,60) = 1.287$, $p = 0.287$, or OSN, $F(3,60) = 0.986$, $p = 0.406$. However, for *speed*, there was a significant main effect of Experiment for both OSO, $F(3,60) = 3.426$, $p = 0.023$, and OSN, $F(3,60) = 5.126$, $p = 0.003$. Consequently, the following OSO and OSN analyses concerns the data pooled across experiments for accuracy, or each experiment individually for speed (Table 3.1).

One-sample t tests (one-tailed) were used to assess whether any performance changes between time points were significantly greater than zero. Consistent with the original study, we observed significant overnight accuracy improvements for OSO [$M = 4.649$, $SD = 15.089$, $t(63) = 2.465$, $p = 0.008$] and OSN [$M = 5.638$, $SD = 15.081$, $t(63) = 2.991$, $p = 0.002$] when data were pooled across experiments. Aside from Experiment 1, improvements in speed were larger (relative to accuracy) and more in keeping with the original study for both OSN and OSO (Table 3.1).

3.2.2.4 Reconsolidation Score

The critical index of a reconsolidation effect was the percentage difference between Old Sequence performance at the Reminder Stage and Test Stage (from herein Reconsolidation

[22]Only the final three trials of the Training and Interference Stages were used because calculating an average across all 12 trials could attenuate the true time-dependent performance changes achieved by the end of these stages (after Walker et al., 2003).

Table 3.1: Overnight scores (speed) for direct replications.

Experiment	Sequence	Overnight Score	SD	$t_{(15)}$	p
1	old	5.93	20.81	1.14	.136
2	old	16.72	14.16	4.72	<.001
3	old	19.72	16.87	4.67	<.001
4	old	26.00	19.90	5.23	<.001
1	new	2.89	12.86	0.90	.191
2	new	15.82	15.66	4.04	<.001
3	new	13.63	9.25	5.89	<.001
4	new	21.71	16.60	5.23	<.001

Score or RS; Figure 3.1, triangles). The Reconsolidation Score fluctuated tightly around zero for both accuracy (Figure 3.3A) and speed (Figure 3.3B) in all four direct replication attempts (Experiments 1–4). Averaged across experiments, mean RS declined by less than 1% for accuracy (compared with a decline of $\sim 57\%$ in the original study), and increased by $\sim 4\%$ for speed (compared with a decline of $\sim 2\%$ in the original study). One-sample t tests (one-tailed) indicated that none of the RS values (Table 3.2) obtained in the direct replications were significantly less than zero.

The inherent limitations of null-hypothesis significance testing constrain the degree to which one can determine the strength of evidence in favor of the null hypothesis (Dienes, 2014). To address this we conducted a Bayesian analysis that enabled us to quantify the evidence in favor of the null hypothesis H_0 (RS $= 0$) relative to the reconsolidation hypothesis H_1 (RS < 0). Specifically, we calculated directional Bayes factors (Rouder et al., 2009) using an 'objective' JZS prior (Cauchy distribution with scale $r = 1$). H_1 was based on the general prediction of reconsolidation theory that trace-dependent performance should be reduced following disrupted reconsolidation of the reactivated trace (Nader & Hardt, 2009; Schwabe et al., 2014; Tronson & Taylor, 2007). In all experiments, Bayes factors (BF_{01}; Table 3.2) were larger than 1, indicating greater evidentiary support for H_0 relative to H_1.

A primary goal of replication attempts is to facilitate more precise estimates of effect-size

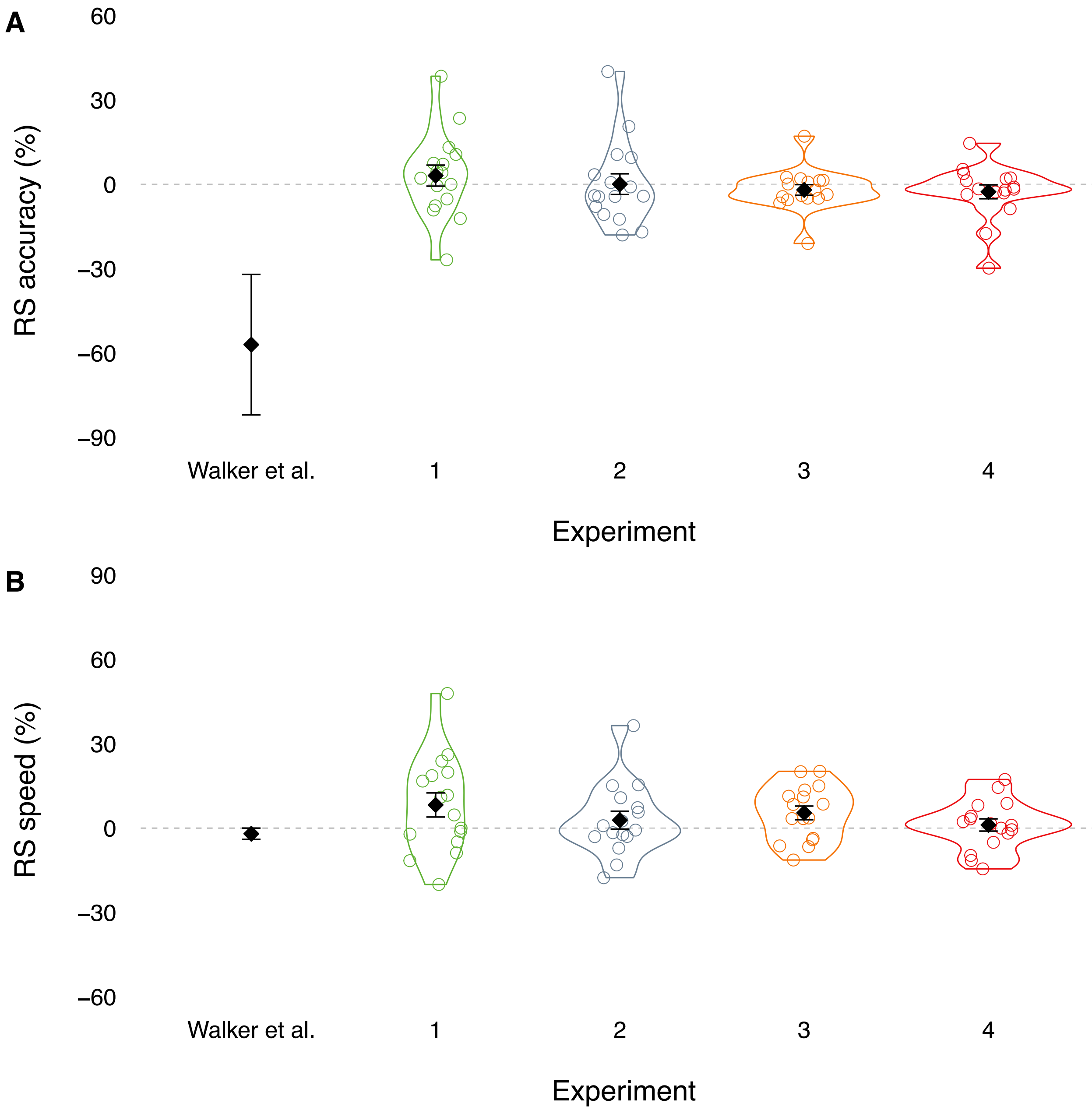


Figure 3.3: Accuracy (**A**) and speed (**B**) reconsolidation scores (RSs) for Walker et al. (2003; $n = 16$) and Experiments 1–4 ($N = 64$). Black diamonds represent means, and error bars show SEM. Where raw data are available (Experiments 1–4), individual participant scores (circles, with random horizontal jitter) and kernel density distributions are also depicted.

Table 3.2: Direct replication RS statistics for accuracy and speed.

Experiment	Dependent Variable	RS	SD	$t_{(15)}$	p	BF_{01}
1	accuracy	3.13	14.94	0.84	.792	8.97
2	accuracy	0.08	14.75	0.02	.508	5.38
3	accuracy	-1.98	7.58	-1.04	.157	1.90
4	accuracy	-2.64	9.88	-1.07	.151	1.85
1	speed	8.24	17.14	1.92	.963	13.59
2	speed	2.85	12.74	0.89	.807	9.23
3	speed	5.41	9.75	2.22	.979	14.68
4	speed	1.12	8.70	0.52	.693	7.52

RS, Reconsolidation Score. BF_{01}, Bayes factor quantifying evidence in favor of the null hypothesis (RS = 0) relative to the reconsolidation hypothesis (RS < 0).

magnitude (Cumming, 2012). However, in light of the stark discrepancy between the finding observed in the original experiment ($n = 16$) and the four direct replications ($n = 64$), we focused on assessing the extent to which the collated evidence indicated that the phenomenon exists at all. That is to say, we aimed to establish whether the effect is qualitatively reproducible, as non-replication will preclude attempts to derive greater quantitative precision in the estimation of the effect's magnitude. Directional meta-analytic Bayes factors using t values for Experiments 1-4 (Table 3.2) indicated greater evidentiary support for the null hypothesis (RS = 0) relative to the reconsolidation hypothesis (RS < 0) for both accuracy ($BF_{01} = 5.743$) and speed ($BF_{01} = 36.027$). This pattern remained after incorporating an estimated t value for the original study[23] (accuracy: $BF_{01} = 2.080$; speed: $BF_{01} = 31.317$).

[23] As neither raw or summary-level data for the original study were available, we used plot-digitizing software to extract RS values for accuracy (M = -57, SEM = 25) and speed (M = -2, SEM = 2) from the original article (Figure 4C in Walker et al., 2003). Values were rounded to the nearest whole number. As t values were not reported, we used these means and SEMs to recalculate them for use in meta-analysis.

3.2.3 Discussion

Consistent with the original study (Walker et al., 2003), we observed time-dependent improvements in accuracy and speed across the course of the Training and Interference Stages, and overnight between stages (Figure 3.2). However, contrary to reconsolidation theory, we saw no marked decline in performance between reactivation and test. The Reconsolidation Scores for accuracy completely contradicted the finding of the original study: we observed only minor fluctuations around zero in all four direction replications. For speed, we also observed only small, non-significant effects (Walker et al., 2003 observed a small, non-significant decline). Bayesian analysis confirmed that there was greater evidentiary support for the null hypothesis compared to the reconsolidation hypothesis for all Reconsolidation Scores. The complete absence of predicted outcomes across these four experiments suggests that the reconsolidation effect reported in Group 7 of the original study (Walker et al., 2003) is not robust.

3.3 Experiments 5-7

In the second component of the replication battery we conducted a series of conceptual replications ($n = 48$) using 'declarative' recall conditions more consistent with the wider human reconsolidation literature (e.g., J. C. K. Chan & LaPaglia, 2013; Forcato et al., 2009; Hupbach et al., 2007). These experiments also involved sequence learning within a 3-day reconsolidation protocol (Figure 3.1), but used sequences similar in length and structure to phone numbers (Experiments 5 and 7) or computer passwords (Experiment 6). A No-Reminder control group (Experiment 7) enabled us to ascertain whether performance impairments were contingent on retrieval-induced plasticity as predicted by reconsolidation theory.

3.3.1 Methods

3.3.1.1 Participants

Sixteen participants were randomly allocated to each of the three conceptual replication experiments, affording a total sample size of 48 individuals (38 females; age Mdn = 22 years, range = 18-52 years). Three additional participants were excluded as they did not complete all three stages of the study.

3.3.1.2 Design

Participants performed a sequence-learning task in three discrete sessions taking place on consecutive days (Figure 3.1). Two 10-item sequences with independent grammars[24] were assigned to be the Old Sequence and the New Sequence in counterbalanced order. For Experiments 5 and 7, the sequences were numbers (X: 1–4–6–3–2–9–5–0–8–7; Y: 2–6–5–7–0–1–9–4–3–8). For Experiment 6, the sequences were letters (X: l–p–k–s–f–q–j–d–x–h; Y: j–f–l–d–q–x–k–h–p–s). On day 1, an adaptive test-feedback protocol was used to ensure that all participants could recall the Old Sequence unassisted five times in a row (Training). On day 2, participants in Experiments 5 and 6 recalled and restudied the Old Sequence immediately before new learning (Reminder). All participants learned the New Sequence in the same manner as Old Sequence Training (Interference). On day 3, participants were asked to recall both sequences in counterbalanced order (Test). The dependent variable was a metric of the similarity between the target (Old/New) sequence at a given stage and the

[24]Sequences were generated with relatively unique grammars but used the same items to ensure a degree of old-new competition. To do this, we first defined a 'base set' of 10 items, which were either randomly selected consonants (l p k s f q j d x h; Experiment 6) or single digits (0–9; Experiments 5 and 7). The first sequence was generated by randomly shuffling the order of these items. The second sequence was generated by repeatedly shuffling the first sequence until (i) all relative item positions (i.e., pairwise forward and backward transitions) were unique, and (ii) all absolute item positions were unique. The same two sequences were used for all participants.

sequence entered by the user ('sequence similarity')[25].

Sequence order (i.e., whether the Old Sequence was designated as 1–4–6–3–2–9–5–0–8–7 / 2–6–5–7–0–1–9–4–3–8 in Experiments 5 and 7; or l–p–k–s–f–q–j–d–x–h / j–f–l–d–q–x–k–h–p–s in Experiment 6; with the remaining sequence being assigned as the New Sequence) and test order (i.e., whether the Old or New Sequence was tested first on the Day 3 Test) were counterbalanced ($n = 8$ per condition). Neither test order nor sequence order influenced the presence/absence of reconsolidation effects (see Appendix B).

3.3.1.3 Procedure

Participants were seated in front of a computer screen in a quiet room and responded using a standard keyboard. On STUDY trials, participants were instructed to memorize the sequence while it was displayed on screen for 5s. No response was required. On $\text{RECALL}_{\text{Feedback}}$ trials, participants were asked to enter the sequence from memory into 10 blank place-holders (_). Correctly entered items appeared in green. Entering an item in an incorrect order caused that item to flash in red and black (4 × 0.5s flashes over 2s) followed by replacement with the correct item, which flashed in green and black (4 × 0.5s flashes over 2s), and early termination of the trial. On $\text{RECALL}_{\text{NoFeedback}}$ trials, participants also had to enter the sequence from memory; however, the trial was not interrupted if they entered items in an incorrect order and they could make corrections if they wished. All items appeared in black so there was no feedback on these trials.

The Training and Interference Stages involved iterative cycles of STUDY and $\text{RECALL}_{\text{Feedback}}$ trials starting with the former. Accurately entering the whole sequence on a $\text{RECALL}_{\text{Feedback}}$ trial led to additional $\text{RECALL}_{\text{Feedback}}$ trials. Failure to complete a $\text{RECALL}_{\text{Feedback}}$ trial resulted in a STUDY trial and the cumulative $\text{RECALL}_{\text{Feedback}}$ counter was reset. When the participant had achieved five accurate $\text{RECALL}_{\text{Feedback}}$ trials in a row, the stage was

[25]Specifically, the similarity between the target (Old/New) and user-entered sequences was measured using a normalized ratio of the Damerau–Levenshtein edit distance: a metric that indicates the number of 'fundamental' operations (substitution, deletion, insertion, or transposition) required to convert one character string into another and thus reflecting the 'similarity' of the two sequences (Van der Loo, 2014).

terminated.

The Reminder Stage involved a single $\text{RECALL}_{\text{NoFeedback}}$ trial followed by two STUDY trials. The Test Stage involved two $\text{RECALL}_{\text{NoFeedback}}$ trials where participants were asked to 'Recall the OLD sequence from day one and enter it on the next screen' and, separately, 'Recall the NEW sequence from day two and enter it on the next screen.' Participants completed sessions at precise 24-h intervals (±15 min); however, session times varied between participants (9:00 AM to 6:00 PM).

3.3.2 Results

Unequal variances in between-subject comparisons were addressed by using Welch t tests. Statistical significance was defined at the .05 level.

3.3.2.1 Training Stage (Old Sequence)

All participants were trained until they reached a performance criterion of five consecutive errorless recalls of the Old Sequence (Figure 3.4). Recall failure resulted in additional study trials. This ensured that, regardless of idiosyncratic learning strategies, all participants robustly encoded the sequence. More trials were required to reach criterion in Experiment 6 (Letters; M = 13.00, SD = 6.623) than Experiment 5 (Numbers; M = 8.688, SD = 2.676) and Experiment 7 (Numbers No Reminder; M = 7.813, SD = 3.124). A one-way ANOVA indicated that the number of trials required to reach criterion varied significantly between experiments, $F(2,45) = 6.089$, $p = 0.005$. Follow-up Welch two-sample t tests indicated that Experiments 5 and 7 did not differ significantly, $t(29.31) = 0.851$, $p = 0.402$. However, participants required significantly more trials to reach criterion in Experiment 6 compared with Experiment 5, $t(19.77) = -2.42$, $p = 0.026$. No participants failed to reach the performance criterion.

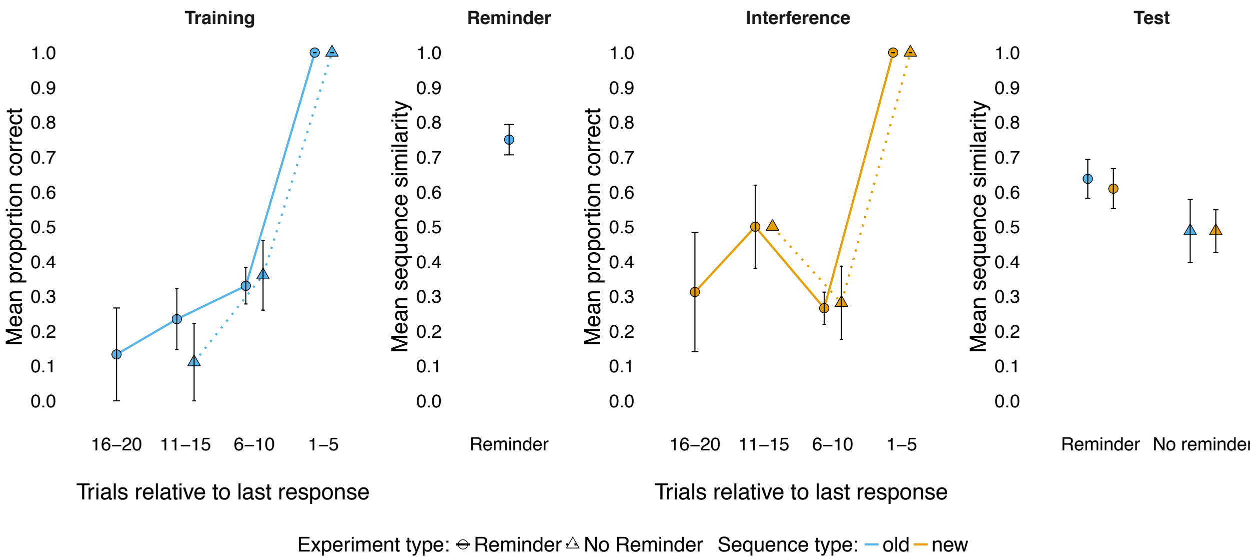


Figure 3.4: Full study timeline showing performance in Experiments 5 and 6 pooled (Reminder groups; n = 32), and Experiment 7 (No-Reminder group; n = 16). The Training and Interference panels show mean proportion correct on $\text{RECALL}_{\text{Feedback}}$ trials across five trial bins plotted relative to participants' final response of the stage. All participants reached the performance criterion (five correct trials in a row) but required a different number of trials to do so. The small number of participants who took more than 20 trials to reach criterion (Training: n = 2, maximum trials = 29; Interference: n = 1, maximum trials = 22) contribute to all relevant analyses. The Reminder and Test panels show mean sequence similarity between the target sequence and the user-entered sequence assessed on a single $\text{RECALL}_{\text{NoFeedback}}$ trial for each previously learned sequence (Old and New). Error bars show SEM.

3.3.2.2 Interference Stage (New Sequence)

The same learn-to-criterion procedure was used in the Interference Stage as in the Training Stage. Although more trials were required to reach criterion in Experiment 6 (Letters; M = 10.375, SD = 4.938) than in Experiment 5 (Numbers; M = 8.563, SD = 4.397) and Experiment 7 (Numbers No Reminder; M = 7.125, SD = 2.363), a one-way ANOVA indicated that these differences were not statistically significant, $F(2,45) = 2.583$, $p = 0.087$. No participants failed to reach the performance criterion.

3.3.2.3 Between-Stage Comparisons

Following the Training Stage baseline (1.0), there were decrements in Old Sequence performance at the subsequent Reminder (M = 0.756, SD = 0.253) and Test Stages (M = 0.606, SD = 0.315) in Experiment 5 (Numbers). A similar pattern was observed in Experiment 6 (Letters) with performance declining at the Reminder (M = 0.744, SD = 0.248) and Test Stages (M = 0.669, SD = 0.322). A 2 × 3 mixed-factorial ANOVA (Experiment: 5, 6; Stage: Training, Reminder, Test) showed that this performance decline across stages was statistically significant, $F(2,87) = 8.629$, $p < 0.001$. There was no main effect of Experiment, $F(1,87) = 1.122$, $p = 0.292$, or interaction between Experiment and Stage, $F(2,87) = 0.672$, $p = 0.513$.

New Sequence performance declined from the Interference Stage baseline (1.0) to the subsequent Test Stage in Experiment 5 (M = 0.625, SD = 0.317), Experiment 6 (M = 0.594, SD = 0.342), and Experiment 7 (M = 0.488, SD = 0.245). A 3 × 2 mixed-factorial ANOVA (Experiment: 5, 6, 7; Stage: Interference, Test) indicated that there was a main effect of Stage, $F(1,88) = 15.425$, $p < 0.001$, and no main effect of Experiment, $F(2,88) = 0.548$, $p = 0.580$, or interaction between Experiment and Stage, $F(2,88) = 0.548$, $p = 0.580$.

3.3.2.4 Reconsolidation Score

As the overall performance pattern did not vary between Experiments 5 and 6 ('Reminder Experiments'), we pooled these data for subsequent analysis. A one-way repeated-measures ANOVA indicated significant changes across stages (Training, Reminder, Test) for the Reminder experiments (see Figure 3.4), $F(2,90) = 8.68$, $p < 0.001$. Follow-up paired t tests (one-tailed) showed that there was significant decline from the Training Stage (1.0) to the Reminder Stage (M = 0.750, SD = 0.246), $t(31) = 5.742$, $p < 0.001$, and from Reminder stage to the Test Stage (M = 0.638, SD = 0.315), $t(31) = 2.645$, $p < 0.001$. Participants in the No-Reminder Control Group also showed a substantial performance decrement from Training (1.0) to Test (M = 0.488, SD = 0.363). A paired-samples t test (two-tailed) confirmed that this decline was significant, $t(15) = 5.646$, $p < 0.001$. A between-group comparison of Test Stage performance indicated poorer recall in the No-Reminder Experiment (M = 0.488, SD

= 0.363) compared to the Reminder Experiments (M = 0.638, SD = 0.315), although a two-sample t test (one-tailed) indicated no significant difference, $t(26.59)$ = -1.409, p = 0.915.

3.3.3 Discussion

During the Training and Interference Stages, all participants successfully reached the criterion of five consecutive errorless sequence recalls indicating successful learning of both the Old and New Sequences. In Experiments 5 and 6 (Reminder), there was a decline in performance between baseline (Reminder Stage) and Test Stage, as anticipated by reconsolidation theory. However, a similar decline was also observed in Experiment 7 (No Reminder control). Therefore, the recall impairments observed in the Reminder Experiments could not be causally attributed to the time-dependent interaction of memory reactivation and interference as required by reconsolidation theory (Nader & Hardt, 2009; Schwabe et al., 2014; Tronson & Taylor, 2007). Indeed, rather than inducing a transient state of heightened susceptibility to interference, memory reactivation resulted in numerically *less* recall impairment of the Old Sequence at the Test Stage relative to no memory reactivation, an effect in the opposite direction to that predicted by reconsolidation theory.

3.4 General Discussion

Reconsolidation-updating theory suggests that retrieval of an existing trace in the human memory system can render that trace vulnerable to modification from post-retrieval new learning. In the present investigation, we attempted to replicate and extend a critical finding (Walker et al., 2003) widely considered to provide a compelling demonstration of reconsolidation-mediated memory updating in humans. In four direct-replication attempts involving procedural recall and three conceptual-replication attempts involving declarative recall, we did not observe the critical impairment effects observed in the original study and predicted by reconsolidation theory (Nader & Hardt, 2009; Schwabe et al., 2014; Tronson & Taylor, 2007).

Broadly speaking, a non-replication of this kind could occur for three reasons: (1) the original finding was a false-positive; (2) the replication was a false-negative, or (3) some unanticipated variable moderated the effect. It is not possible to definitively determine which of these explanations is correct, but we can consider the weight of evidence for the various possibilities. Naturally the larger sample size obtained in the present study relative to the original study affords greater confidence in the interpretation that the replications were not false-negatives, and that the original study could have been a false-positive. Is there additional data bearing on this question? Several recent investigations have used variations of the original paradigm (Walker et al., 2003), with either transcranial magnetic stimulation (TMS, Censor et al., 2010, 2014b, 2014a) or new learning (Beukelaar et al., 2014) interventions, but there is some contention as to whether these studies successfully replicate the findings of the original study or not (cf. Hardwicke & Shanks, 2016; Walker & Stickgold, 2016). Although the findings of these studies were interpreted as yielding favorable evidence for reconsolidation theory, the expected Reminder–Test performance decrement was actually absent in most conditions.

The study by Beukelaar et al. (2014) represents a reasonably close procedural replication of the original and present study, although the primary dependent variable was operationalised in a different manner. Specifically, ‘accuracy’ was the percentage of correct sequences on a given trial and ‘speed’ was the inter-trial time interval measured in seconds (cf. Appendix A). The primary ‘performance’ measure was a ratio of accuracy/speed, and between stage comparisons were based on shorter ‘blocks’ (first 5 tapped sequences). In contrast to Walker et al. (2003), and consistent with the present findings, no Day 2 - Day 3 performance decrement was observed in a group who performed 3x30s finger tapping trials during the Reminder Stage. A number of additional groups were given different reminder configurations (5x sequences, 10x sequences, 1x30s, 1x60s), and Beukelaar et al. (2014) did observe a significant linear trend in which shorter reactivation lengths were associated with greater Day 2 - Day 3 decrements. This is an interesting pattern that may warrant further empirical attention. However, in order to test a reconsolidation account it will be necessary to demonstrate that this effect is due to a reminder by intervention interaction (see Chapter 2) rather than just an effect of the reminder manipulation itself. It should also be noted that the observed decrements were modest and performance rapidly recovered within the test session. It is difficult to reconcile this with the prediction of permanent trace modification (Nader & Hardt, 2009).

In the series of experiments that employed TMS (Censor et al., 2010, 2014b, 2014a) the researchers focused on the 'speed' dependent measure as operationalised in Walker et al. (2003) and the present replications. The critical accuracy measure that showed a substantial decrement in the original study was not examined. On the speed measure, there was actually a small Day 2 to Day 3 performance increment for the R^+I^+ group, and these gains were less than that seen in a R^+I^- group. This effect was interpreted as favourable to reconsolidation theory because the 'offline gains' achieved in the R^+I^- group were assumed to have been mediated by a reconsolidation process. The reduced gains in the R^+I^+ group were therefore taken to reflect 'blocked offline gains' as a result of TMS disrupting the reconsolidation process.

But there are some important caveats to bear in mind regarding these results. Firstly, it should be noted that this is a qualitatively different effect to the one reported in Walker et al. (2003), which was a substantial *decrement* in performance (accuracy) between Day 2 and Day 3. Furthermore, the effect is rather tenuous evidence for reconsolidation: (i) there was no "no-reminder" control condition (R^-I^+ or R^-I^-; see C4, C6, C7 in Chapter 2); (ii) the intervention was delivered during, rather than after reactivation (see C2 in Chapter 2); and (iii) such "offline gains" can often be driven by various procedural confounds (Rickard, Cai, Rieth, Jones, & Ard, 2008). In summary, the outcomes of this set of studies (Beukelaar et al., 2014; Censor et al., 2010, 2014b, 2014a) cannot be straightforwardly interpreted as yielding successful replications of the original finding (cf. Hardwicke & Shanks, 2016; Walker & Stickgold, 2016), and hardly seem consistent with the central tenets of reconsolidation theory more generally (see Chapter 2, Nader & Hardt, 2009; Tronson & Taylor, 2007).

As noted above, another possible explanation for the discrepancy in findings is that some unanticipated variable(s) moderated the effect. In other words, it could be the case that reconsolidation theory in its present form is not sufficiently well specified to identify the conditions under which reliable effects will and will not be observed. The authors of the original study have recently proposed several such moderators (Walker & Stickgold, 2016), which include participant age range, and session time, and suggest that our experiments have established these potential moderators as 'boundary conditions' on reconsolidation theory.

However, this claim blurs the important line between the exploratory ("hypothesis-generating") and confirmatory ("hypothesis-testing") phases of scientific inquiry (Wagenmakers et al., 2012). Post-hoc conjecture, or "hypothesing after the results are known" (Kerr, 1998) is an important part of the scientific process as it can generate new hypotheses. However, because it is easy to 'over-fit' explanations once the data have been observed, such conjectures should be considered tentative until they have been empirically verified (Hardwicke & Shanks, 2016).

Furthermore, the role of the specific moderators that Walker and Stickgold (2016) suggest can be examined to some extent through additional analyses of our data (see section B.1). In the case of age range for instance, it is possible to re-examine the critical Reconsolidation Scores in the subgroup of participants who fell within the exact age bracket of participants recruited in the original study (18-27). This still yields a healthy sample size of 48 participants, and the additional analyses confirm that no reconsolidation effect emerged. Similarly, analyses indicated no appreciable relationship between Reconsolidation Scores and session time. These analyses suggest that the proposed moderators do not provide a compelling account of the extant data. Nevertheless, our experiment was not designed to manipulate these variables as they were not indicated as relevant in the original study (Walker et al., 2003), nor are they considered central tenets of reconsolidation theory (Nader & Hardt, 2009). It is important that proposed 'boundary conditions' are not only applied on a case-by-case basis to instances of theoretical failure, but are also able to accommodate instances of theoretical success. Thus, if these potential moderators are to be considered boundary conditions then they should ideally maintain or extend, rather than restrict, the explanatory coverage of the theory (see section 2.3).

The findings of our conceptual replications cast further doubt on the veracity of claims that memory updating can be mediated by reconsolidation processes (J. C. K. Chan & LaPaglia, 2013; Dudai, 2009; Hardt et al., 2010; J. L. Lee, 2009; Schiller et al., 2010). These experiments adhered to the canonical 3-day reconsolidation protocol and aimed to increase external validity through the use of sequences similar in structure to phone numbers or computer passwords. In addition, consistent with several studies in the human reconsolidation literature (e.g., J. C. K. Chan & LaPaglia, 2013; Forcato et al., 2009; Hupbach et al., 2007), participants completed a declarative recall task. Under these conditions, performance impairments occurred in the both the presence and absence of memory retrieval. Rather than triggering a state of heightened trace-vulnerability, retrieval actually led to numerically higher performance compared to the no-retrieval control group. This finding is consistent with previous investigations of reconsolidation that found retrieval practice can afford some protection against interference (e.g., Potts & Shanks, 2012), and a considerable body of evidence suggesting that retrieval aids rather than impairs subsequent recall (Roediger & Butler, 2011).

Two notable aspects of human reconsolidation research are not directly addressed by the present investigation. First, there is evidence that post-retrieval pharmacological interventions can attenuate emotional responding in a fear-conditioning paradigm (Kindt et al., 2009). However, the reliability of these effects has also recently come under scrutiny (Bos, Beckers, & Kindt, 2014). Similarly, initially promising findings based on using post-retrieval extinction to disrupt reconsolidation in a fear-conditioning paradigm (Schiller et al., 2010) have proved elusive in subsequent replication attempts (Golkar et al., 2012; Kindt & Soeter, 2013). It is striking that declarative recall of the conditioned stimulus–unconditioned stimulus contingency remains intact across these fear-conditioning studies, either in the presence or absence of effects on emotional responding.

Additionally, it has been suggested that 'prediction error' is a necessary reconsolidation trigger (see Chapter 2, Forcato et al., 2009; Sevenster et al., 2014). If this were the case, it could explain the absence of reconsolidation effects in the present replications. However, it would also be surprising that a reconsolidation effect was observed in the original study because the reminder protocol required participants to practice the Old Sequence in its entirety, and thus presumably did not invoke prediction error. To justify an auxiliary theoretical assumption about prediction error, one would need to reconcile a considerable amount of contradictory evidence. For example, relative to controls, no impairment of declarative recall is observed in the aforementioned prediction error studies (Forcato et al., 2009; Sevenster et al., 2014), only attenuation of emotional responding (Sevenster et al., 2014), or ambiguous null effects on an indirect measure of trace integrity (retrieval-induced forgetting; Forcato et al., 2009). Furthermore, reconsolidation-like effects have been reported when the reminder involves reinforced trials (and thus no prediction error) in both non-human animals (Duvarci & Nader, 2004) and humans (J. C. K. Chan & LaPaglia, 2013), and impairment effects are absent even in studies where prediction error would be expected (Golkar et al., 2012; Kindt & Soeter, 2013). At present, therefore, it is unclear whether prediction error is either necessary or sufficient for reconsolidation effects to emerge.

Taken together, our findings cast doubt on the efficacy of new-learning interventions as a means for disrupting the reconsolidation of procedural or declarative memory in humans. The absence of reconsolidation effects in all four direct replications suggests that the considerable theoretical weight attributed to the original study (J. L. Lee, 2009; Nader & Hardt, 2009; Schiller & Phelps, 2011; Schwabe et al., 2014; Tronson & Taylor, 2007) is no longer warranted. Furthermore, the absence of retrieval-contingent impairment in the conceptual replications is inconsistent with the purported functional role of reconsolidation as an adaptive mechanism that underlies memory updating (J. C. K. Chan & LaPaglia, 2013; Dudai, 2009; Hardt et al., 2010; J. L. Lee, 2009). Replication will be an essential tool in future reconsolidation investigations as researchers seek to verify the reliability of existing findings, identify genuine boundary conditions, and foster theoretical progress.

Chapter 4

Revisiting the misinformation effect: Does disruption of reconsolidation enable memory trace overwriting?

The classic 'misinformation effect' powerfully testifies to the inherent fallibility of human memory in general and the potential unreliability of eyewitness testimony in particular (Ayers & Reder, 1998; Loftus, 2005; Zaragoza et al., 2006). In a seminal study by Loftus and colleagues (Loftus et al., 1978), participants first viewed a series of slides depicting a traffic accident (the Event Phase). Subsequently, one group of 'misled' participants was exposed to information that contradicted an aspect of the event information whereas for another group of 'control' participants no misinformation was presented (the Post-Event Information Phase). For example, when a 'stop' sign was shown in the slides, the post-event information either referred to a 'stop' sign (control condition[26]) or a 'yield' sign (misled condition). In the Final Test Phase, all participants were asked what happened during the event and presented with a forced choice between the event item and the misleading item. 75% of participants in the control group correctly chose the event item compared to just 41% of misled participants. Exposure to misinformation had apparently led a considerable number of individuals to report something that never actually happened during the original event.

Whilst these striking findings highlight the ease with which eyewitness testimony can potentially be distorted (Zaragoza et al., 2006), their mechanistic origins have been the subject of considerable controversy. As outlined in Chapter 1, the misinformation effect is a specific example of the broader phenomenon of 'retroactive interference', whereby learning of new information impairs an individual's ability to recall previously learned information (Crowder,

[26] In the control condition of Loftus et al. (1978) the event item was correctly represented at the Post-Event Information Stage (e.g., referring to a 'stop sign' as a 'stop sign'). However, in subsequent studies (e.g., Loftus, Donders, Hoffman, & Schooler, 1989) the typical approach has been to make a 'neutral reference' to the event item (e.g., referring to a 'stop sign' as a 'sign').

1976; Titcomb & Reyna, 1995; Postman & Underwood, 1973). A historic debate spanning multiple fields of enquiry, experimental paradigms, and species, has considered whether these apparent 'forgetting' effects are attributable to permanent alteration of underlying memory traces (i.e., "storage-based" impairment, e.g., J. M. Barnes & Underwood, 1959; Müller & Pilzecker, 1900; Melton & Irwin, 1940) or factors that may only temporarily perturb trace-dependent performance while allowing the trace itself to persist (i.e., "retrieval-based" impairment, e.g., Bouton, 2002; Capaldi & Neath, 1995; J. A. McGeoch, 1942; Tulving, 1974). The processes that drive retroactive interference may also have an adaptive role through inhibiting and/or replacing erroneous or out-dated information stored in the memory system. Therefore, identifying the underlying cause of the misinformation effect will not only aid efforts to mitigate the influence of misleading information on eyewitness testimony, it will impact upon our broader understanding of the fundamental operations that enable human memory updating (R. A. Bjork, 1978; Dudai, 2009).

An early account of the misinformation effect suggested that exposure to misleading information resulted in the 'destructive updating' or 'overwriting' of the memory trace representing the event information (Loftus, 1979a, 1979b; Loftus et al., 1985, 1989). In other words, not only were participants led to preferentially report misinformation, they had also permanently erased aspects of the event information from memory. However, as outlined in Chapter 1 (also see Figure 1.1), poor performance on a memory test does not necessarily entail any destructive loss of information from the memory system. Such an effect could indeed arise when target information is no longer available (i.e., permanently destroyed), but it would also occur when memory traces are temporarily inaccessible (e.g., Tulving, 1974). In the years following Loftus's initial investigations, a lively debate ensued about the type of memory impairment that might cause the misinformation effect (see Figure 4.1). When retrieval factors were implicated (e.g., Bekerian & Bowers, 1983; Chandler, 1991; Christiaansen & Ochalek, 1983), researchers suggested that the trace representing the event information was intact (the 'persistence hypothesis'). When retrieval factors were not identified, researchers often favoured storage-based accounts such as the 'overwriting hypothesis' (e.g., Belli, Windschitl, & McCarthy, 1992; Chandler, 1989; Loftus, 1979b). The theoretical and practical implications of the two accounts are stark. The overwriting hypothesis entails the irrecoverable loss of

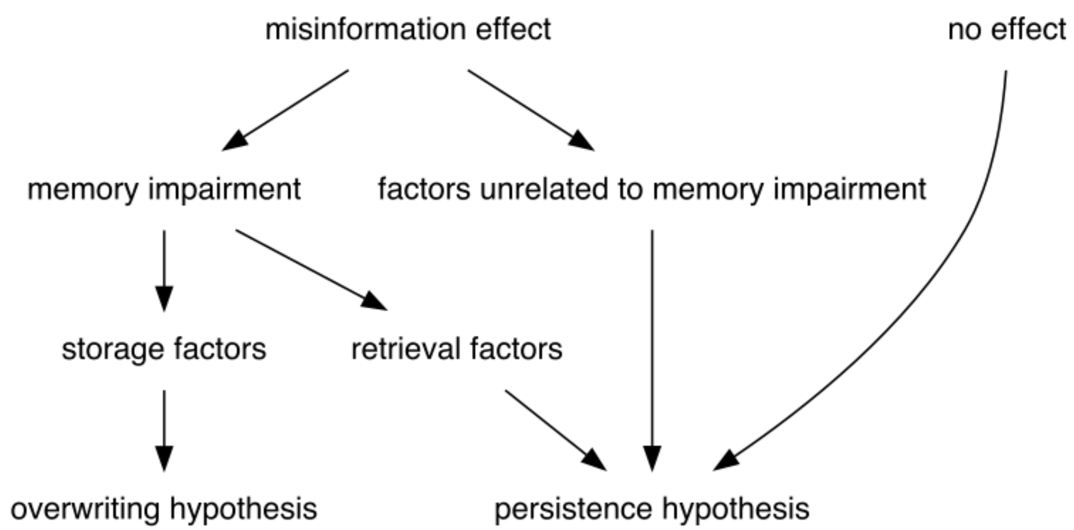


Figure 4.1: Theoretical implications of the misinformation effect (control performance > misled performance). If a misinformation effect is observed using conventional tests (e.g., the Original Test or the True/False Test) then it could be due to memory impairment or factors unrelated to memory impairment. When a misinformation effect is observed using a test that minimizes the role of factors unrelated to memory impairment (e.g., the Modified Test), it suggests that memory impairment has occurred. However, the locus of the memory impairment (i.e., storage/retrieval factors) remains to be determined. Obtaining a misinformation effect using the Modified Test is therefore necessary, but not sufficient evidence in support of the overwriting hypothesis. If no misinformation effect is observed (control performance ≈ misled performance) then this corroborates the persistence hypothesis.

memory traces whereas the persistence hypothesis suggests that those traces are intact and potentially accessible, particularly under favourable retrieval circumstances (Chandler & Fisher, 1996).

In considering the relative merits of these memory impairment accounts, it is important to note that a participant's performance on a memory test can be influenced by a number of factors unrelated to the presence or retrievability of information stored in memory, such as motivation, attention, or fatigue (Cahill et al., 2001). Therefore, an additional type of explanation that does not involve memory impairment at all is also plausible (Figure 4.1). In a rigorous critique, McCloskey and Zaragoza (1985a, also see 1985b; Zaragoza & McCloskey, 1989) outlined how the memory test employed in previous investigations (from herein "Original

Test"; e.g., Loftus et al., 1978) was susceptible to confounding factors. Recall that on the Original Test participants are offered a two-alternative forced choice between the event item and the misleading item. McCloskey and Zaragoza (1985a) suggested that poorer performance in the misled condition relative to the control condition is expected even if no memory impairment occurs. Firstly, consider a participant who may not remember the event item during the Final Test Phase for reasons unrelated to the misleading information, for example, never encoding the event item during the Event Phase[27]. In the control condition, such a participant should have no particular inclination towards choosing either the event item or the misled item[28] and would simply be guessing. However, in the misled condition, a participant who does not remember the event item may nevertheless be able to remember the misled item, and would therefore be biased towards choosing it ("misinformation acceptance", also see Belli, 1989). Secondly, consider a participant who recalls both the event item and the narrative reference. In the control condition, such a participant would likely choose the event item, because it was not contradicted during the Post-Event Information Phase. However, in the misled condition, the participant may be more inclined to choose the misled item by reasoning that they are likely mistaken: presumably the experimenter who provided the narrative knew what happened in the slides better than they do ("deliberation", also see Belli, 1989). Alternatively, the participant might choose the misled item believing it to be the response that the experimenter wants them to give ("demand characteristics", also see Lindsay, 1990). Finally, a participant who can remember one or both items may simply respond based on item familiarity and neglect information about the item's source entirely ("source misattribution"; also see Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989). Overall then, there are a number of reasons why the inclusion of the post-event item as an option on the Original Test will lead to poorer performance in the misled condition compared to the control condition for *reasons unrelated to memory impairment*.

McCloskey and Zaragoza (1985a) reinforced their arguments by devising a new 'Modified Test'.

[27]This is likely to be the case for many participants as performance in the control condition is typically well-below ceiling (Loftus et al., 1978; Zaragoza et al., 2006).

[28]Any item-level effects, such as the general plausibility of an item's role in the scenario, are minimized through counter-balanced assignment of items to the control and misled conditions.

The procedure is similar to the Original Test, requiring participants to make a two-alternative forced choice between the event item and a second item. However, instead of the misled item, the second option is a novel item that has never occurred during either the Event Phase or the Post-Event Information Phase. The absence of the post-event item from the test should minimize the impact of response biases in the misled condition, thus providing a less confounded assessment of memory impairment (see Figure 4.1). If misinformation hinders an individual's ability to remember the event information (through either storage-based or retrieval-based memory impairment) then participants will have no reason to prefer the event item to the novel item and should perform at chance level in the misled condition. By contrast, if participants can successfully recall the event information then performance in the misled condition should be similar to the control condition. Across six experiments ($N = 444$) using the Modified Test[29], McCloskey and Zaragoza (1985a) found that performance in the control and misled conditions was indeed comparable (75% vs. 72%, *ns*[30]). The authors concluded that there was no evidence that exposure to misleading information had any impact on participants' ability to remember the event information, thus corroborating the persistence hypothesis.

Since McCloskey and Zaragoza (1985a), a number of other investigations have employed the Modified Test to evaluate whether misinformation effects can be attributed to memory impairment. Payne, Toglia, and Anastasi (1994) identified a total of 44 such experiments (reported in 13 individual studies) of which the majority (30) did not report a statistically significant difference between the control and misled conditions. However, the authors noted that when the outcomes of all studies were combined in a single overall t-test, there was a

[29]Note that in the study described at the outset of this article (Loftus et al., 1978), post-event item type (misled vs. control) was manipulated between-subjects. Subsequent studies (e.g., J. C. K. Chan & LaPaglia, 2013; Loftus et al., 1989; M. McCloskey & Zaragoza, 1985a) have employed a within-subjects manipulation where several event items were subsequently referred to in a misleading way, or a neutral way (control), during the Post-Event Information Phase.

[30]Although such null findings are ambiguous within the framework of null-hypothesis significance testing (e.g., Dienes, 2014), a Bayes Factor can be calculated using the reported t value and sample size (see "General Analysis Procedure" for more details). The Bayes Factor ($BF_{01} = 3.866$) indicates moderate support for the null hypothesis (control - misled = 0) relative to the directional alternative hypothesis (control - misled > 0).

small but significant difference between the control ($M = 75.8\%$) and misled ($M = 71.7\%$) conditions, $t(43) = 4.65$, $p < .001$. Contrary to McCloskey and Zaragoza (1985a), they concluded that memory impairment can contribute to the misinformation effect. Although we will raise some potential problems with this analysis in the discussion, Payne et al. (1994)'s study appears to have shifted the focus of the misinformation debate from asking *Does memory impairment ever occur?*, to *Under what circumstances does memory impairment occur?* (Ayers & Reder, 1998; Loftus, 2005).

Recently, it has been proposed that storage-based impairment effects can arise when participants are exposed to contradictory misinformation during a transient period of retrieval-induced plasticity initiated during memory 'reconsolidation' (J. C. K. Chan & LaPaglia, 2013; Hardt et al., 2010). Reconsolidation theory suggests that existing memory traces are destabilized when they are reactivated (e.g., via retrieval) and start to undergo a time-dependent restabilisation (reconsolidation) process. During this brief time window, it is suggested that traces are temporarily amenable to modification, or even erasure, if the reconsolidation process is disrupted (J. C. K. Chan & LaPaglia, 2013; Nader, 2003a; Schiller et al., 2010). Reconsolidation effects are investigated in non-human animals through the delivery of invasive interventions intended to perturb the molecular restabilisation of reactivated memory traces (Nader & Hardt, 2009; Tronson & Taylor, 2007). For example, in a fear-conditioning paradigm Nader, Schafe, and Le Doux (2000a) observed the apparent loss of a previously acquired freezing response in rats following post-retrieval infusion of a protein-synthesis inhibitor, anisomycin, directly into the animals' amygdala. By contrast, the freezing response was intact in no-reactivation, no-intervention, and delayed-intervention control groups, indicating that the performance deficit was contingent on the predicted time-dependent interaction of reactivation and intervention.

It is not straightforward to investigate reconsolidation in humans primarily because the invasive treatments used in non-human animal studies have toxic side effects (Agren, 2014; Schiller & Phelps, 2011; cf. Kroes et al., 2014). However, a number of investigators have used post-retrieval new learning as a means to 'update' or 'overwrite' existing traces by disrupting the reconsolidation process (e.g., Hupbach et al., 2007; Forcato et al., 2007; Schiller et al., 2010; St Jacques & Schacter, 2013; Walker et al., 2003). Recently, it has been specifically proposed that disruption of reconsolidation could enable the overwriting of event memory traces in a misinformation paradigm (J. C. K. Chan & LaPaglia, 2013; Hardt et al., 2010; also see J. C. K. Chan et al., 2009). In Chan and LaPaglia (2013), participants watched a video episode of the TV series '24' during the Event Phase. In the Post-Event Information Phase, an audio narrative was presented in which some of the critical items in the video were referred to in a misleading way ('misled' condition), or were not mentioned ('control' condition). For example, in the video, a fictional terrorist used a hypodermic needle on a flight attendant and in the misled condition the audio narrative suggested that the terrorist had used a stun gun. In the Final Test Phase, statements about critical items from the video were presented, some of which were true ('the terrorist used a hypodermic needle') and some of which were false ('the terrorist used a chloroform rag'). As with the Modified Test, the misleading item was omitted: the false statements referred to a novel item that had not appeared in either the video or the audio narrative. For each statement, participants responded either 'true' if they had seen the item in the video or 'false' if they had not (from herein 'True/False Test').

In order to investigate the role of reconsolidation, Chan and LaPaglia (2013) introduced a Reminder Phase that preceded the Post-Event Information Phase. During the Reminder Phase, one group of participants ('Reactivation' condition) answered a series of cued-recall questions about the critical items and other items from the video. The goal was to reactivate event memory traces, initiating their reconsolidation and rendering them temporarily vulnerable to overwriting. Another group of participants ('No Reactivation' condition) played the game Tetris for 10 minutes as a distractor activity. An important feature of the canonical reconsolidation protocol is the relative timing of the experimental phases. Although Chan and LaPaglia (2013) reported four experiments in which they predicted impairment effects, only one of these (Experiment$_{\mathrm{CL}}$ 6) adhered to the timing parameters required for a controlled test of reconsolidation theory (e.g., Nader et al., 2000a; also see Chapter 2; Nader & Hardt, 2009; Tronson & Taylor, 2007; Schiller & Phelps, 2011). In Experiment$_{\mathrm{CL}}$ 6, the phases were distributed across 3 consecutive days (see Figure 4.2). This ensured that (a) there was time for information encoded during the Event Phase on Day 1 to consolidate overnight before (b) the Reminder Phase immediately preceded the Post-Event Information Phase on Day 2, and

(c) the Final Test Phase did not occur until Day 3 when impairment effects attributable to the disruption of reconsolidation are expected to emerge (Nader et al., 2000a; Nader & Hardt, 2009). The other experiments deviated from key aspects of this design (specifically points *a* and *c*) either because all phases took place on the same day (Experiment$_{CL}$ 1, Experiment$_{CL}$ 4), or the Final Test Phase was conducted immediately after the Reminder Phase on Day 2 (Experiment$_{CL}$ 3). Thus, only Experiment$_{CL}$ 6 was capable of ruling out the key confounds that are controlled in a three-day design. Consistent with the predictions of reconsolidation theory, performance in Experiment$_{CL}$ 6 was impaired in the misled condition compared to the control condition only in the Reactivation group, and not in the No-Reactivation group[31]. Chan and LaPaglia (2013) concluded that "human declarative memory can be selectively rewritten during reconsolidation" (p. 9309) and noted that theirs was the first study to demonstrate reconsolidation-associated impairment effects on a declarative recall measure.

Are the findings observed in Experiment$_{CL}$ 6 evidence of overwriting? As outlined above (Figure 4.1), corroboration of the overwriting-hypothesis would first require unambiguous evidence that the misinformation effect was due to memory impairment before one could begin to unravel the relative contributions of storage and retrieval factors. However, as with the Original Test, the True/False Test used in Chan and LaPaglia (2013) may also be influenced by factors unrelated to memory impairment. Indeed, the True/False Test[32] was originally intended only to minimize the influence of misinformation acceptance (Belli, 1989). It has been acknowledged that the test cannot straightforwardly delineate the source of any residual performance deficits, which could still be attributed to memory impairment (storage- or retrieval-based), source misattribution, and/or deliberation (Belli, 1989; Lindsay & Johnson,

[31] Although this pattern of results appears consistent with reconsolidation theory, no direct comparison of the Reactivation/No-Reactivation conditions is reported (see C4 and C7 in Chapter 2). This is potentially problematic because the difference between significant and not significant is not necessarily itself statistically significant (Gelman & Stern, 2006; Nieuwenhuis et al., 2011). Additionally, the apparent absence of an effect in the No-Reactivation group is difficult to reconcile with classic misinformation studies that did not include a reactivation stage but still observed performance deficits using the True/False Test (Belli, 1989; Lindsay & Johnson, 1989; Tversky & Tuchin, 1989).

[32] Rather than requiring a true/false response, early versions of the test required an equivalent 'yes/no' response (Belli, 1989; Lindsay & Johnson, 1989; Tversky & Tuchin, 1989).

1989; Zaragoza & McCloskey, 1989). Lindsay and Johnson (1989), for instance, found that misinformation effects observed on a True/False Test were eliminated when participants were explicitly directed to consider item source (also see Zaragoza & Lane, 1994; Zaragoza & Koshmider, 1989).

Therefore, as with the Original Test, performance on the True/False Test in the misled condition could be lower than in the control condition for reasons unrelated to memory impairment. As it is necessary to demonstrate memory impairment before one can identify its locus (storage or retrieval factors, see Figure 4.1) the True/False Test may not be optimal for evaluating whether disruption of reconsolidation enables the overwriting of event memory traces. In summary, the observation of poorer performance in the misled vs. control condition in Chan and LaPaglia's (2013) $\text{Experiment}_{\text{CL}}$ 6 (Reactivation Condition) cannot be unambiguously interpreted as evidence in favour of the reconsolidation-overwriting account.

In the present study, we developed a 3-day hybrid reconsolidation-misinformation paradigm and sought to test the prediction that reactivation prior to the presentation of contradictory misinformation would facilitate the overwriting of memory traces representing a witnessed event (J. C. K. Chan & LaPaglia, 2013; Hardt et al., 2010). We used the Modified Test (M. McCloskey & Zaragoza, 1985a) as the principle means of assessing participants' event knowledge to enhance our confidence that any observed performance deficits were not contaminated by factors unrelated to memory impairment. Detecting a misinformation effect on the Modified Test would be necessary but not sufficient evidence in support of the overwriting hypothesis. Further investigation would be required to delineate the relative contributions of storage-based and retrieval-based factors (see Figure 4.1). The absence of a misinformation effect on the other hand, would demonstrate that event memory traces remained intact despite exposure to contradictory misinformation during the putative reconsolidation window. This would be inconsistent with the reconsolidation-overwriting hypothesis, and corroborate the persistence hypothesis.

Alongside Modified Test outcomes, we sought to verify that our novel hybrid paradigm could elicit previously reported misinformation effects with the Original Test (Experiment 8) and the True/False Test (Experiment 9). As previous studies have used different methods for triggering 'reconsolidation', we also investigated whether specific reactivation procedures are necessary to elicit memory impairment effects (Experiment 10). Finally, we revisit and re-evaluate the findings of a previous meta-analysis (Payne et al., 1994) that suggested misinformation effects can be obtained using the Modified Test, and thus attributed to memory impairment.

4.1 General Methods

All data, and analysis code pertaining to this chapter have been made publicly available on the Open Science Framework (https://osf.io/7wn8c). All data exclusions, manipulations, and measures conducted during this study are reported.

4.1.1 Overview

All of the experiments reported here employed the three critical phases of the classic 'misinformation paradigm', plus an additional phase designed to trigger memory reconsolidation (see Figure 4.2). As in Experiment$_{\mathrm{CL}}$ 6, the phases were distributed across three consecutive days in order to meet the timing parameters of the canonical reconsolidation paradigm (Nader & Hardt, 2009; Tronson & Taylor, 2007). On Day 1, participants watched a short video in order to learn the initial 'event' information (*Event Phase*). The following day (Day 2), participants were reminded about items that appeared in the event (*Reminder Phase*) immediately prior to the *Post-Event Information Phase*, during which they read a narrative containing references to critical events from the video. On the final day (Day 3), participants completed a forced-choice recognition test containing items that appeared during the event video and foil items (*Final Test Phase*). The core design involved a within-subjects manipulation at the Post-Event Information Phase: four critical items were referred to in a misleading manner ('Misled' condition), and four critical items were referred to in a neutral manner ('Control' condition).

In the experiments outlined below, the Event Phase and Post-Event Information Phase did not vary, however we manipulated the nature of the Final Test Phase (Original Test, True/False Test, Modified Test) and the Reminder Phase (No Reminder, Subtle Reminder, Retrieval Practice) in order to investigate various predictions of the reconsolidation-overwriting hypothesis. The overall goal was to evaluate the claim that reminding participants about the event video would trigger the reconsolidation of target memory traces and render them vulnerable to 'overwriting' by subsequently presented misinformation.

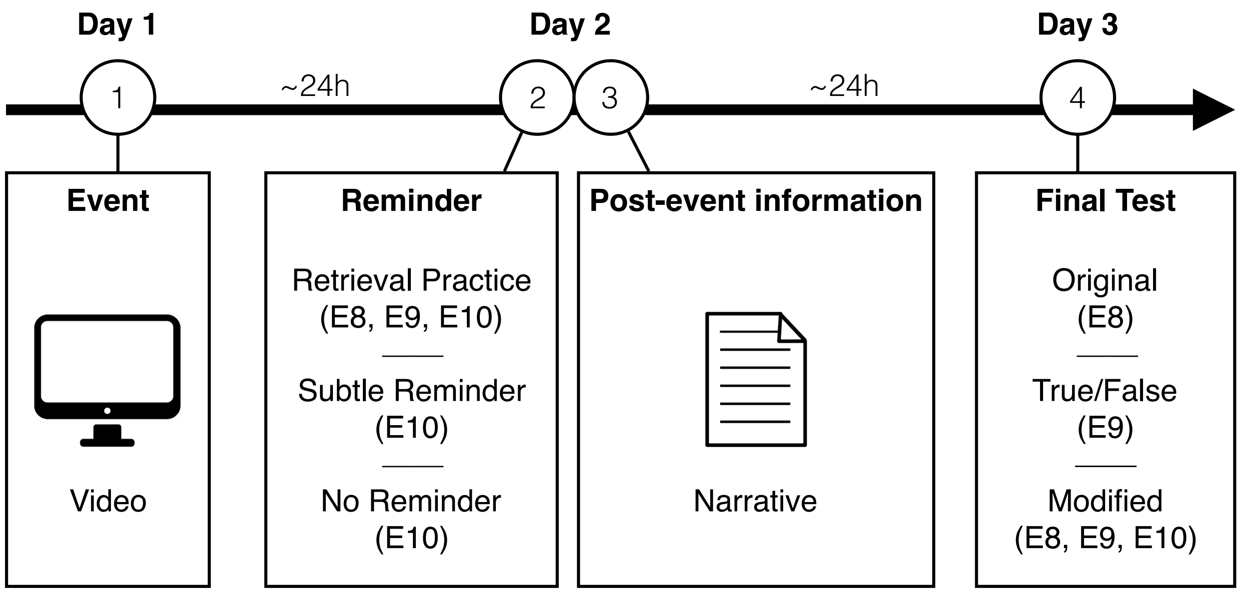


Figure 4.2: Hybrid reconsolidation-misinformation paradigm used in Experiments 8-10. The four discrete phases of each experiment were distributed over 3 consecutive days. The nature of the Reminder Phase and Final Test Phase varied across experiments whereas the Event Phase and Post-Event Information Phase remained constant.

4.1.2 Materials

Modern video (6m 28s) and narrative (803 words) materials developed specifically for the misinformation paradigm (Takarangi, Parker, & Garry, 2006) were employed. The video depicted a tradesman, 'Eric', visiting an unoccupied home and examining the owner's belongings as he made his way through various rooms. Eric helps himself to food and drink and steals a number of items such as jewelry. Both the video and narrative contained 12 'filler' items (see Appendix D) included to disguise the aims of the experiment, and 8 critical items (see Appendix C) that were subsequently referred to in the post-event narrative. Four of the critical items were referred to in a misleading way ('Misled' condition), and four of the critical items were referred to in a neutral way ('Control' condition). For example, a participant who saw Eric drink a can of Pepsi in the video might subsequently read that he 'helped himself to a *can of Coke* from the fridge' (Misled item; emphasis added) or 'helped himself to a *can of soft drink* from the fridge' (Control item; emphasis added). Two versions of the video and four versions of the narrative were employed in order to counter-balance[33] which type of each critical item appeared in the video and whether it was subsequently referred to in a misleading or neutral manner (see Appendix E). For example, under optimal counter-balancing conditions half of the participants would see Coke during the video and half would see Pepsi. In addition, half of the participants would read a misleading reference to the drink in the narrative (either Coke or Pepsi, depending on the item presented in the video) and half would read a neutral reference.

4.1.3 General procedure

The experiments were delivered to participants via a web browser using custom JavaScript code written by T.E.H. Participants were told that they were participating in a study investigating 'visual and verbal learning modes' (after Takarangi et al., 2006), instructed to rely only on their memory, and asked not to write down any information.

[33]It should be noted that the structure of the materials did not enable us to include the 'novel item' within the counter-balancing arrangements. If the novel item is to serve as an effective foil option on the Modified Test, then it is important that there is no bias towards selecting either the event item or the novel item (for a participant who has no knowledge of either). In other words, it is necessary to ensure that a given item is not inherently more plausible, or more similar to the misled item, than its alternative on the Modified Test. One way to examine the influence of this potential confound would be to ask a new group of participants to read a version of the narrative that contains only neutral references, and then ask them to complete the Modified Test. As these participants will not have been previously exposed to either the event or novel item, their performance should be at chance level if they are not inherently biased towards either option. Ultimately, it will be necessary to replicate the present experiments using materials that allow for complete counter-balancing of the event, misled, and novel items.

4.1.3.1 Event Phase

On Day 1, participants were told to pay close attention to the video, as they would be asked questions about it later in the study. After watching the video, they were asked to return after a 24-hour interval for the next stage of the study.

4.1.3.2 Reminder Phase

On Day 2, participants in the Retrieval Practice condition (Experiments 8-10) answered 20 randomly intermixed and sequentially presented cued-recall questions about the critical items (see Appendix C) and filler items (see Appendix D). For example, "What did Eric drink a can of?" The questions were self-paced and a response was required before participants could move on to the next item. For each response, participants also used a slider to provide a confidence rating from 0-100%. Participants in the Subtle Reminder condition (Experiment 10) were simply told, "Please take a few moments to think about the video you saw yesterday" and presented with a 30 second countdown after which the phase terminated. Participants in the No Reminder Condition (Experiment 10) skipped this phase entirely.

4.1.3.3 Post-Event Information Phase

Immediately after the Reminder Phase, participants were instructed to read the narrative once only. Reading time was self-paced.

4.1.3.4 Final Test Phase

On Day 3, participants were told, "During this stage we will ask you a series of questions about the video you saw during stage 1. We are testing your memory for this video." Participants then completed the Original Test (Experiment 8), True/False Test (Experiment 9), or Modified Test (Experiments 8-10). For all tests, questions regarding the 8 critical items (see Appendix C) and 12 filler items (see Appendix D) were presented in a random order. Items were self-paced and a response was required before participants could move on to the next item. For each response, participants also used a slider to provide a confidence rating from 0-100%.

For the Original Test, participants were presented with questions about the items and given two response options: the event item and the misled item (see Appendix C and Appendix D). The relative screen position (top/bottom) of each item type was randomised. Participants were instructed to choose the item that most accurately represented what they saw in the video (see Appendix F).

For the True/False Test, participants were presented with statements about the items (see Appendix C and Appendix D). Half of the statements were correct and referred to the event item. Half of the statements were incorrect and referred to a novel item that had not appeared in either the video or the narrative. The number of correct/incorrect statements was balanced across Post-Event Information conditions such that two of each type of item (misled and control) were referred to incorrectly (i.e., statement about novel item) and two were referred to correctly (i.e., statement about event item). None of the statements referred to the misled item. Participants were instructed to respond 'true' if they saw a statement that accurately represented what they saw in the video, or 'false' if it did not (see Appendix F).

The Modified Test was identical to the Original Test except that the response options were the event item and a novel item that had not appeared in either the video or narrative. The misled item was never an option.

4.1.4 General analysis procedure

All data used in ANOVAs met the assumption of homogeneity of variance (according to Levene's test). Where data used in *t* tests did not meet the assumption of normality (according to a Shapiro-Wilk test), we also employed an equivalent non-parametric test (e.g., Wilcoxon Signed-Rank). Unless otherwise reported, the outcome of non-parametric tests were consistent with the parametric tests. Median values for all key pairwise comparisons are also shown in Figure 4.3.

4.1.4.1 Final Test Phase Accuracy

On the Original and Modified Tests, item accuracy (i.e., choosing the 'event item') was first aggregated at the participant level providing a 'proportion of correct responses' score for both misled and control items. On the True/False Test, response accuracy was first calculated at the participant level as the hit rate minus the false alarm rate for both misled and control items.

For the critical effect under scrutiny (δ) we considered the null hypothesis (H_0) of no difference in accuracy between control and misled conditions ($\delta = 0$), relative to the directional alternative hypothesis (H_1) of higher accuracy in the control condition compared to the misled condition ($\delta > 0$), as predicted by reconsolidation theory (see introduction). Support for H_1 on the Modified Test would indicate memory impairment and therefore be *consistent with* the overwriting of the underlying memory trace representing the event (see Figure 4.1). Further investigation would be required to delineate the exact nature of the impairment effect (storage-based and/or retrieval-based). Support for H_0 would demonstrate that overwriting had not occurred and corroborate the persistence hypothesis.

Our analysis approach consisted of a complementary battery of classical null hypothesis significance testing (NHST), Bayesian statistics, and cumulative effect size estimation through meta-analysis (G. S. Howard, Maxwell, & Fleming, 2000). For each critical comparison, statistical significance for NHST was defined at the .05 level.

Because of the interpretative difficulties surrounding null effects within a NHST framework (Dienes, 2014), the conclusion of 'no impairment' in previous Modified Test studies has always been necessarily tentative (Zaragoza, Dahlgren, & Muench, 1992). To overcome this inferential asymmetry, we calculated directional Bayes Factors (BF_{01}; Rouder et al., 2009) using a 'default' JZS prior (Cauchy distribution with scale $r = .707$). Bayes Factors can be understood as a continuous measure of evidentiary strength for one hypothesis (e.g., H_0) relative to a second hypothesis (e.g., H_1) and are therefore capable of indicating support for the null. For instance, $BF_{01} = 3$ would indicate that the observed data are three times more likely under the null hypothesis relative to the alternative hypothesis.

4.1.4.2 Final Test Phase Confidence

The difference between confidence ratings (0-100%[34]) in the control vs. misled condition was also examined. We first calculated mean confidence ratings for both misled and control items at the participant level. We considered the null hypothesis (H_0) of no difference between conditions ($\delta = 0$), and the non-directional[35] alternative hypothesis (H_1) of differential accuracy in the control condition compared to the misled condition ($\delta <> 0$).

4.1.4.3 Retrieval Practice Condition Coding

For individuals in the Retrieval Practice condition it was possible to gain some insight into participants' event knowledge at the Reminder Phase. This also enabled an analysis of responses in the Final Test Phase contingent on recall accuracy during the Reminder Phase. This is useful because only when items have been successfully retrieved during the Reminder Phase can we be relatively certain that the underlying traces (a) were initially formed and remained intact immediately prior to the Post Event Information Phase; and (b) were successfully reactivated, and thus presumably underwent reconsolidation.

As responses during Retrieval Practice were unconstrained, manual coding was necessary to quantify accuracy. A response was coded as 'correct' when its meaning approximately matched the event information. For example, for the question "What was Eric wearing on the lower half of his body?" (see Appendix C), the imperfect response 'blue trousers' would be considered sufficiently similar to the correct response "jeans", and marked as correct. As there was a degree of subjectivity in this process, two raters independently coded each item. Inter-rater reliability was high (Cohen's $\kappa = 0.92$) and the small number of discrepancies were resolved through discussion.

[34]Participants in Experiment 8 actually provided confidence ratings on a scale ranging from 0-5. However, these responses have been converted to percentages so as to be consistent with confidence ratings provided in Experiments 9 and 10, which were on a 0-100% scale.

[35]Although previous studies have found higher confidence ratings for control vs. misled items, we employed non-directional tests as an effect in the opposite direction remains plausible. These analyses can be considered relatively exploratory as the *a priori* predictions of reconsolidation theory are not clear in this context.

4.1.4.4 Filler items

Performance on filler items was high and comparable across experiments (M = 0.90, SD = 0.01) indicating good general retention of event details. No further analysis was conducted on filler performance as it was not relevant to the hypotheses under scrutiny.

4.2 Experiment 8

This experiment had two principle goals: (1) to establish whether our hybrid reconsolidation-misinformation paradigm could replicate the classic misinformation effect when participants' knowledge was measured using the Original Test (e.g., Loftus et al., 1978); and (2) to evaluate whether memory impairment effects could be detected using the Modified Test under conditions that should induce a transient state of plasticity in target memory traces whilst they undergo reconsolidation (J. C. K. Chan & LaPaglia, 2013; Hardt et al., 2010). As outlined in the introduction, the Modified Test is a more appropriate assay of memory impairment because it minimizes the influence of unrelated factors such as misinformation acceptance, deliberation, demand characteristics, and source-misattribution (M. McCloskey & Zaragoza, 1985a).

As in Chan and LaPaglia (2013), participants engaged in 'retrieval practice' during the Reminder Phase in order to trigger reconsolidation of the memory traces representing event information. During the Final Test Phase one group of participants completed the Original Test and another group completed the Modified Test.

4.2.1 Methods

4.2.1.1 Participants

Fifty-nine first-year psychology undergraduate students (52 female; 7 male; age Mdn = 18 years, range = 18-22 years) attending University College London (UCL) were recruited as part of a class on research methods. Participants were randomly assigned to the final test conditions (Original Test, n = 35; Modified Test, n = 24). 52 additional participants attempted the study but were rejected prior to data analysis due to failure to comply with the necessary timing parameters (n = 32), technical issues that resulted in data loss (n = 7), or not attending to the video or narrative[36] (n = 13). These data exclusions resulted in unbalanced assignment to counter-balancing conditions for experimental materials (A = 19, B = 24, C = 7, D = 9; see Appendix E), an issue that is addressed in Experiment 10. In this and all subsequent experiments, participants provided informed consent and received either monetary compensation or course credits. The local UCL ethics committee approved the study.

4.2.1.2 Design and procedure

The experiment was a 2 x 2 mixed-factorial design with Post-Event Information (Control, Misled) manipulated within-subjects and Final Test (Original, Modified) manipulated between-subjects. The primary dependent variable was the mean proportion of correct responses (i.e., choosing the 'event' item) on either the Modified Test or the Original Test.

The Event Phase (Day 1) and Final Test Phase (Day 3) were completed remotely on participants' personal computing devices. For the Reminder Phase and Post-Event Information Phase (Day 2) participants were tested under close supervision in three batches of approximately equal size in a large university computer room. Participants were instructed to maintain as close to a 24-hour interval between daily sessions as they could manage (time between Day 1 end and Day 2 start: M = 22.13h, SD = 4.08h; time between Day 2 end and Day 3 start: M = 25.08h, SD = 3.67h).

4.2.2 Results and discussion

A 2 x 2 mixed-factorial ANOVA with Post-Event Information (Misled, Control) and Final Test (Original, Modified) as independent variables and mean proportion correct as the dependent

[36] Participants were strictly instructed to watch the video and read the narrative in a single sitting. Our software monitored whether these rules were adhered to, for example by recording use of the video 'pause' button.

variable indicated a significant main effect of Post-Event Information, $F(1,57) = 43.72$, $p < .001$, $\eta^2_G = .27$, a significant main effect of Final Test, $F(1,57) = 77.10$, $p < .001$, $\eta^2_G = .42$, and a significant interaction between the two factors, $F(1,57) = 31.63$, $p < .001$, $\eta^2_G = .21$. We therefore proceeded with pairwise comparisons between control and misled conditions independently for the Original Test and the Modified Test (see Figure 4.3).

The misinformation effect observed in the Original Test group was consistent with previous studies (e.g., Loftus et al., 1989, 1978). For misled items, the proportion of correct responses was substantially lower than performance for control items, and even fell below chance level (50%), demonstrating that participants were strongly inclined to report misinformation rather than event information. Furthermore, consistent with previous studies (Loftus et al., 1989; Takarangi et al., 2006), responses in the misled condition were endorsed with higher confidence ratings compared to responses in the control condition (see Table 4.1).

However, there was an unexpected difference in baseline performance during the Reminder Phase (Table 4.2). Specifically, retrieval practice performance was lower for items that were subsequently selected for the misled condition compared to items that were subsequently selected for the control condition. Although this difference was not statistically significant in the parametric analysis, a Shapiro-Wilk test suggested that the assumption of normality could be breached ($W = 0.90$, $p = .004$), and a Wilcoxon-Signed Rank (non-parametric) test was statistically significant ($z = 5.78$, $p = <.001$). Nevertheless, when the data were analysed contingent on correct responding during Reminder Phase retrieval practice, overall accuracy was higher, and a large difference between control and misled conditions remained (Table 4.3). Therefore, the observed misinformation effect cannot be solely attributed to differences in baseline performance between the control and misled conditions.

According to the reconsolidation-overwriting hypothesis (J. C. K. Chan & LaPaglia, 2013; Hardt et al., 2010) the misinformation effect observed in the Original Test group is at least partially driven by the erasure of event memory traces. However, despite invoking conditions that should disrupt memory reconsolidation, performance in the Modified Test group indicated that the event memory traces had persisted (Figure 4.3). Consistent with previous studies (e.g., M. McCloskey & Zaragoza, 1985a), no difference was observed between performance in the control and misled conditions when the Modified Test was employed. To avoid inferential ambiguity in the interpretation of this null effect we employed Bayes Factors which indicated moderate evidence ($BF_{01} = 5.102$) in favour of the null hypothesis relative to the reconsolidation-overwriting hypothesis. Interestingly, Modified Test responses were endorsed with higher confidence ratings for control items compared to misled items suggesting that misinformation did have some impact on participants' strength of belief in the veracity of event item responses, despite not influencing their recognition accuracy (see Table 4.1).

In the Modified Test group, Reminder Phase performance was comparable for the control and misled conditions (Table 4.2). When Final Test performance was analysed contingent on Reminder Phase performance overall performance was elevated, but there was still no difference between control and misled conditions (Table 4.3). Therefore, the absence of a misinformation effect cannot be attributed to a failure to reactivate the memory traces representing the event information (a necessary trigger for reconsolidation).

In summary, Experiment 8 has demonstrated that misinformation effects detected using the Original Test are not indicative of reconsolidation-mediated overwriting of event memory traces. Responses in the Modified Test reveal that those traces persist, even under conditions that should disrupt their reconsolidation.

4.3 Experiment 9

The findings of Experiment 8 are inconsistent with the theory that reconsolidation enables the overwriting of memory traces in the misinformation paradigm (J. C. K. Chan & LaPaglia, 2013; Hardt et al., 2010). However, it has previously been suggested that performance impairments observed on a True/False Test were indicative of a reconsolidation-overwriting effect (J. C. K. Chan & LaPaglia, 2013). In order to examine this claim in more detail, our second experiment aimed to (1) to replicate Chan and LaPaglia's (2013) finding of a misinformation effect with the True/False Test; and (2) to ask the *same participants* to respond to the *same items* using the Modified Test to examine whether any performance deficits detected on the True/False Test could be attributed to overwriting. Even if the findings of Chan and

Test / Exp	Reminder	Control M (SD)	Misled M (SD)	*t*	*p*	BF_{01}
Original Test						
8	Practice	.65(.24)	.21(.20)	8.28	<.001	5e−08
True/False Test						
9	Practice	.48(.41)	.35(.47)	1.54	.064	1.284
Modified Test						
8	Practice	.79(.23)	.80(.18)	−0.12	.549	5.102
9	Practice	.79(.21)	.75(.26)	0.86	.198	3.248
10	Practice	.82(.23)	.85(.20)	−0.56	.710	8.960
10	Subtle	.80(.20)	.82(.20)	−0.43	.667	8.302
10	None	.84(.20)	.86(.17)	−0.53	.701	8.810
Meta–analysis (Modified Test +Reminder Only)						
				Random effects model		12.082

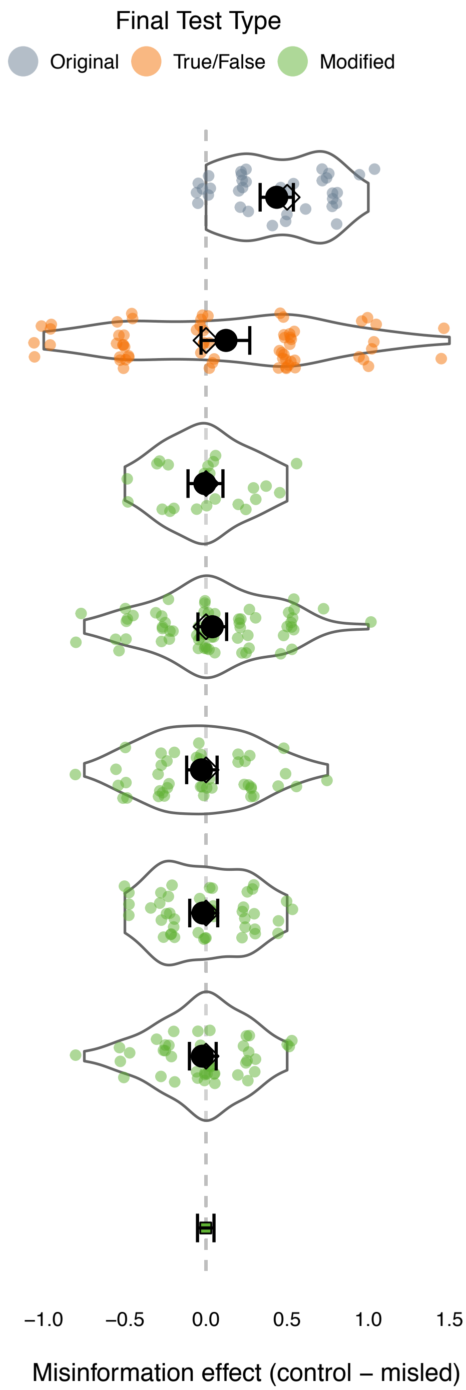


Figure 4.3: Descriptive and inferential statistics for Experiments 8-10. Figure shows means (black circles), 95% CIs (error bars), medians (white diamonds, often overlapping with means), individual participant scores (small coloured circles), kernel density distributions (grey lines) and a combined effect size (square) estimated by a random effects model for Reminder-only (Practice or Subtle) Modified Test experiments.

Table 4.1: Descriptive and inferential statistics for confidence ratings (0-100%) provided during the Final Test Phase.

Exp	Test	Reminder	Control $M(SD)$	Misled $M(SD)$	M_{diff} [95%CI]	t	p	BF_{01}
1	Orig	Practice	64.57(18.45)	76.00(20.32)	-11.43[-18.42,-4.44]	-3.32	.002	0.062
1	Mod	Practice	64.38(16.11)	56.04(20.38)	8.33[0.08,16.59]	2.09	.048	0.744
2	T/F	Practice	68.59(18.43)	76.23(17.77)	-7.64[-12.74,-2.54]	-2.99	.004	0.130
2	Mod	Practice	65.84(21.48)	67.22(23.16)	-1.38[-7.22,4.45]	-0.47	.637	6.601
3	Mod	Practice	65.58(20.79)	61.43(25.18)	4.16[-3.36,11.67]	1.12	.271	3.428
3	Mod	Subtle	62.03(18.17)	61.55(20.46)	0.48[-7.46,8.41]	0.12	.904	6.083
3	Mod	None	61.12(25.08)	63.77(21.83)	-2.65[-9.51,4.20]	-0.78	.439	4.599

Exp, Experiment. Test types: *Orig* = Original, *T/F* = True/False, *Mod* = Modified. BF_{01}, Bayes factor quantifying evidence in favor of the null hypothesis (H_0) relative to the reconsolidation hypothesis (H_1).

Table 4.2: Descriptive and inferential statistics for performance during Retrieval Practice.

Exp	Test	Control $M(SD)$	Misled $M(SD)$	M_{diff} [95%CI]	t	p	BF_{01}
1	Orig	.43(.25)	.34(.24)	0.09[-0.02,0.19]	1.61	.058	0.913
1	Mod	.41(.23)	.44(.17)	-0.03[-0.14,0.08]	-0.57	.713	6.794
2	Mod and T/F	.45(.26)	.47(.25)	-0.02[-0.09,0.06]	-0.41	.658	9.841
3	Mod	.43(.29)	.50(.27)	-0.07[-0.18,0.03]	-1.43	.920	13.833

Exp, Experiment. Test types: *Orig* = Original, *Mod and T/F* = Modified and True/False, *Mod* = Modified. BF_{01}, Bayes factor quantifying evidence in favor of the null hypothesis (H_0) relative to the reconsolidation hypothesis (H_1).

Table 4.3: Descriptive and inferential statistics for Final Test performance contingent on correct responses during Retrieval Practice.

Exp	Test	n	Control $M(SD)$	Misled $M(SD)$	M_{diff} [95%CI]	t	p	BF_{01}
1	Orig	24	.93(.16)	.47(.45)	0.46[0.25,0.67]	4.50	<.001	0.006
1	Mod	21	.92(.24)	.93(.18)	-0.01[-0.16,0.13]	-0.17	.567	4.338
2	T/F	17	.91(.20)	.68(.43)	0.24[-0.01,0.48]	2.06	.028	0.745
2	Mod	57	.95(.19)	.89(.28)	0.06[-0.03,0.15]	1.32	.096	3.032
3	Mod	33	.87(.27)	.94(.15)	-0.07[-0.18,0.04]	-1.25	.890	2.636

Exp, Experiment. Test types: *Orig* = Original, *T/F* = True/False, *Mod* = Modified. BF_{01}, Bayes factor quantifying evidence in favor of the null hypothesis (H_0) relative to the reconsolidation hypothesis (H_1).

LaPaglia (2013) are replicated here, the concurrent absence of performance deficits on the Modified Test would suggest that the effect observed in Experiment$_{\text{CL}}$ 6 is not indicative of reconsolidation-overwriting and is more likely attributable to other factors (Belli, 1989; Lindsay & Johnson, 1989; Zaragoza & McCloskey, 1989).

To reduce the extent to which responses on one test might influence responses on the other (e.g., attempting to provide consistent responses on both tests), we counter-balanced the order of test type and introduced an inter-test 'spot-the-difference' filler task (~ 3 minutes). As in Experiment 8 (and J. C. K. Chan & LaPaglia, 2013) all participants engaged in 'Retrieval Practice' during the Reminder Phase in order to trigger memory reconsolidation.

4.3.1 Methods

4.3.1.1 Participants

Sixty-five first-year UCL psychology undergraduate students (52 female; 11 male; 2 unstated; *Mdn* age = 18 years, age range = 18-21 years) were recruited as part of a class on research methods (a different cohort to Experiment 8). 48 additional participants attempted the study but were rejected prior to data analysis due to failure to comply with the necessary timing parameters ($n = 36$), technical issues that resulted in data loss ($n = 8$), or not attending to the video or narrative ($n = 4$). These data exclusions resulted in unbalanced assignment to counter-balancing conditions for experimental materials (A = 13, B = 14, C = 20, D = 18; see Appendix E), an issue that is addressed in Experiment 10.

4.3.1.2 Design and Procedure

The experiment was a within-subjects design with one independent variable: Post-Event Information (Control, Misled). All participants completed two types of Final Test: the Modified Test and the True/False Test. The order of the tests was counter-balanced across participants (True/False then Modified, $n = 32$; Modified then True/False, $n = 33$). Two dependent variables were measured: (1) the mean proportion of correct responses on the Modified Test; and (2) the mean hit rate – false alarm rate on the True/False Test (see subsection 4.1.4 for details). The Event Phase (Day 1) and Final Test Phase (Day 3) were completed remotely on participants' personal computing devices. For the Reminder Phase and Post-Event Information Phase (Day 2) participants were tested under close supervision in three batches of approximately equal size in a large university computer room. Participants were instructed to maintain as close to a 24-hour interval between daily sessions as they could manage (time between end of Day 1 session and start of Day 2 session: $M = 21.36$h, $SD = 3.01$h; time between end of Day 2 session and start of Day 3 session: $M = 25.64$h, $SD = 3.29$h).

4.3.2 Results and discussion

We first employed 2 (Test Order: modified-statements, statements-modified) x 2 (Post-Event Information: misled, control) mixed factorial ANOVAs to identify any influence of Test Order. We did this separately for the two dependent variables: mean proportion correct on the Modified Test and hit rate – false rate on the True/False Test. For Modified Test performance there was no significant main effect of Test Order, $F(1, 63) = 0.03$, $p = .854$, $\eta^2_G = .00$, or interaction with Post Event Information, $F(1, 63) = 0.45$, $p = .506$, $\eta^2_G = .00$. For True/False

Test performance the main effect of Test Order, $F(1, 63) = 3.82$, $p = .055$, $\eta^2_G = .03$, was close to the alpha level ($\alpha = .05$). However, there was no significant interaction of Test Order with Post Event Information, $F(1, 63) = 0.55$, $p = .459$, $\eta^2_G = .00$. Therefore, we collapsed the remaining analyses across Test Order conditions and proceeded with pairwise comparisons between control and misled conditions (see Figure 4.3).

The misinformation effect observed on the True/False Test was relatively small and only 21 out of 65 participants showed performance deficits larger than zero. The difference between conditions was not statistically significant. As in the Original Test group of Experiment 8, responses were endorsed with significantly higher confidence ratings for control items compared to misled items (see Table 4.1).

Although small, the observed effect appears to be numerically consistent with the outcomes of Experiment$_{\text{CL}}$ 6 and thus constitutes an approximate replication of J. C. K. Chan & LaPaglia (2013) original findings. However, despite having almost twice the sample size of the original experiment, a Bayesian analysis indicated that the True/False outcomes observed in Experiment 9 are inconclusive ($BF_{01} = 1.284$, equivalently the data are only 0.78 times more likely under H_1 relative to H_0). A Bayes Factor for Experiment$_{\text{CL}}$ 6 (calculated using statistics reported in the original article: $t = 2.04$, $n = 33$) indicates that the evidence in the original experiment was also relatively inconclusive ($BF_{01} = 0.444$, equivalently the data were only 2.25 times more likely under H_1 relative to H_0). A meta-analytic Bayes Factor combining the data of the two experiments ($n = 98$) indicates moderate support for the alternative hypothesis ($BF_{01} = 0.278$, equivalently the data are 3.59 times more likely under H_1 relative to H_0).

Despite the small effect detected on the True/False test, there was no significant difference between control and misled conditions when the same participants responded to the same items on the Modified Test. This is is consistent with the findings of Experiment 8, and demonstrates that memory impairment, and thus reconsolidation-mediated overwriting, is not a likely cause of the performance deficits obtained using the True/False test (also see Lindsay & Johnson, 1989; Zaragoza & McCloskey, 1989; Belli, 1989). Confidence ratings also did not differ significantly between the control and misled conditions (see Table 4.1).

Retrieval practice performance at the Reminder Phase (which applies to both Final Test types) was comparable for the control and misled conditions (Table 4.2). When analysed contingent on Reminder Phase performance, overall Final Test performance was elevated, but maintained the same qualitative pattern for both test types (Table 4.3). Therefore, the findings cannot be attributed to differences in baseline performance, or a failure to reactivate event memory traces.

4.4 Experiment 10

Experiments 8 and 9 demonstrated that misinformation effects could be detected in a hybrid reconsolidation-misinformation paradigm using either the Original Test or the True/False Test. However, both experiments also showed that misinformation effects were not detected when the Modified Test was used. This is inconsistent with the predictions of the reconsolidation-overwriting hypothesis and suggests that responses on the Original and True/False Tests are influenced by factors unrelated to memory impairment.

In Experiment 10, we introduced a novel manipulation to expand upon the findings of Experiments 8 and 9. The precise nature of the reminder phase was adjusted across three groups as it has been suggested that certain types of reminder can be more or less effective at triggering reconsolidation (see Chapter 2, Forcato et al., 2009; W. C. Gordon, 1983; Hupbach et al., 2008; Sevenster et al., 2014). As in Experiments 8 and 9, one group of participants engaged in retrieval practice at the Reminder Phase. Retrieval practice has several practical benefits (also see J. C. K. Chan & LaPaglia, 2013). For instance, it enables the researcher to check whether participants actually recalled the critical items, thus providing an important manipulation check as to whether (a) the critical items have been successfully encoded and could be successfully retrieved; and (b) that the relevant memory traces have been 'reactivated', and thus, presumably, started to undergo reconsolidation.

However, it has been suggested that reminder trials involving overt rehearsal of information by the participant, or re-presentation of information by the researcher, are not appropriate in reconsolidation studies because they (a) provide an opportunity for relearning that could strengthen the relevant trace (W. C. Gordon, 1983); and (b) do not involve a mismatch between the retrieval cue and the environment (prediction error) which has been proposed as a critical reconsolidation trigger (Forcato et al., 2009; Sevenster et al., 2014). In several human reconsolidation studies a 'subtle reminder' that does not involve overt rehearsal of specific event details has been employed which may avoid these issues (e.g., Hupbach et al., 2007, 2008). A disadvantage of using a subtle-reminder is that one cannot ascertain whether participants have retrieved the relevant information. This is important because simply providing a cue is not always effective at eliciting retrieval (Tulving, 1974), making it unclear whether the underlying traces have actually been reactivated. The theoretical relevance of prediction error also remains unclear because it does not always seem to be a necessary factor to elicit reported reconsolidation effects (e.g., Duvarci & Nader, 2004; St Jacques & Schacter, 2013, see Chapter 2 for further discussion). Nevertheless, we included a second 'Subtle Reminder' group who were simply told, "Please take a few moments to think about the video you saw yesterday" and presented with a 30 second countdown after which the Reminder Phase terminated. To assess whether a reminder had any impact on performance we also included a third 'No Reminder' control group who did not complete the Reminder Phase at all.

Experiment 10 also afforded an opportunity to conduct a rigorous replication of Experiments 8 and 9 (Modified Test only) whilst exerting tighter control over the spatial and temporal context. Participants who failed to follow instructions (e.g., pausing the video) or did not follow precise timing parameters across the three-day study were immediately excluded and replaced to ensure balanced numbers across conditions. A subgroup participated under tightly controlled laboratory conditions: they attended the same testing room at the same time on each day of the study, and always interacted with the same experimenter (see Hupbach et al., 2008). A larger sample took part via the Internet but still adhered to strict timing parameters.

4.4.1 Methods

4.4.1.1 Participants

One-hundred and thirty-two participants (65 female; 64 male; 3 unstated; *Mdn* age = 25 years; age range = 19-50 years) were recruited from the UCL mixed-occupation subject pool (n = 36) or online crowd-sourcing platform Prolific Academic (n = 96). Participants were randomly and evenly assigned to the three reminder conditions (n = 44 per condition) and four counter-balancing conditions (n = 33 per condition). During testing, 100 additional participants were immediately and automatically rejected by the experimental software due to failure to comply with the necessary timing parameters (n = 74), or not attending to the video/narrative (n = 26).

4.4.1.2 Design and procedure

The experiment was a 2 x 3 mixed-factorial design with Post-Event Information (Control, Misled) manipulated within-subjects and Reminder Condition (No Reminder, Subtle Reminder, Retrieval Practice) manipulated between-subjects. As in Experiments 8 and 9, the dependent variable was the mean proportion of correct responses (i.e., choosing the 'event' item) on the Modified Test.

All phases took place across three consecutive days (Figure 4.2). The sample recruited from UCL attended our testing rooms in person and the online sample completed the study remotely on their own personal computing devices.

Participants were instructed to begin each daily session at 24-hour intervals (±2 hours) and our software enforced these timings by immediately ejecting those who missed this time window. This afforded tightly controlled time intervals in both the laboratory group (time between Day 1 end and Day 2 start: M = 23.78, SD = .25h; time between Day 2 end and Day 3 start: M = 23.93h, SD = .35h) and the online group (time between Day 1 end and Day 2 start: M = 23.49, SD = .98h; time between Day 2 end and Day 3 start: M = 23.80h, SD = .81h).

4.4.2 Results and discussion

We first computed a 2 (Post-Event Information: control, misled) x 3 (Reminder Condition: No Reminder, Subtle Reminder, Retrieval Practice) ANOVA with mean proportion correct as the dependent variable. There was no significant main effect of Post-Event Information, $F(1, 129) = 0.78$, $p = .380$, $\eta^2_G = .00$, no significant main effect of Reminder Condition, $F(2, 129) = 1.19$, $p = .308$, $\eta^2_G = .01$, and no significant interaction between the two factors, $F(2, 129) = 0.01$, $p = .989$, $\eta^2_G = .00$. Despite the absence of any statistically significant effects, descriptive and inferential statistics for pairwise contrasts between control and misled conditions are presented in Figure 4.3 to aid comparison with other experiments.

Consistent with the findings of Experiments 8 and 9, no misinformation effects were detected using the Modified Test either when there was No Reminder, a Subtle Reminder, or Retrieval Practice. Furthermore, when reminder phase performance was taken into account (Retrieval Practice condition only), overall performance was elevated but remained comparable in the control and misled conditions (Table 4.3). There were no inter-condition differences in baseline performance during the Reminder Phase (Table 4.2). As in Experiment 9, there was no significant difference in confidence ratings between the control and misled conditions (Table 4.1). Bayesian analysis indicated moderate support for H_0 relative to H_1 in all three Reminder Conditions (Figure 4.3). These findings replicate the outcomes of Experiments 8 and 9 with a larger sample size, and tighter control of spatial and temporal context. They also indicate that the absence of a misinformation effect in those experiments cannot be attributed to the use of a reminder that involves overt retrieval practice.

4.5 Meta-analysis

Across three experiments we have observed clear evidence that even under conditions that should enable reconsolidation-mediated overwriting of event memory traces, it is the parameters of the test-retrieval situation that determine the presence or absence of misinformation effects. Although performance on the Original Test and True/False Test might initially suggest memory impairment, the absence of misinformation effects on the Modified Test reveals that, after being misled, participants are capable of recognizing the event information to the same extent as when they are not misled.

To support this conclusion, we combined Modified Test outcomes across Experiments 8-10 (excluding the Experiment 10 No Reminder group where a reconsolidation effect was not predicted) using both conventional and Bayesian meta-analytic approaches (Figure 4.3). A random effects model ($k = 4$) yielded a combined effect size of -0.002, 95% CI[-0.053; 0.049] indicating at most a negligible effect of misleading information on recall accuracy compared to neutral information.

For the Bayesian meta-analysis we used t values to compute a meta-analytic Bayes Factor for each test type (Morey & Rouder, 2015). As throughout this article, we compared the null hypothesis (H_0) of no difference in accuracy between control and misled conditions ($\delta = 0$), to the directional alternative hypothesis (H_1) of higher accuracy in the control condition compared to the misled condition ($\delta > 0$; see subsection 4.1.4 for details). This analysis indicated strong support for H_0 relative to H_1 ($BF_{01} = 12.082$).

4.6 General discussion

It is well established that learning new information can influence an individual's ability to recall previously learned information (Crowder, 1976). However, these 'retroactive interference' effects can often be attributed to retrieval-based factors and do not necessarily imply that the underlying memory trace has been altered (see Chapter 1, Bouton, 2002; Tulving, 1974). Identifying the locus of interference that causes misinformation effects has proved to be the subject of considerable controversy and various accounts have been proposed that entail storage-based memory impairment (Chandler, 1989; Loftus et al., 1985), retrieval-based memory impairment (Bekerian & Bowers, 1983; Chandler, 1991), or no memory impairment at all (M. McCloskey & Zaragoza, 1985a).

The present study was motivated by renewed interested in the storage-based 'destructive updating' or 'overwriting' hypothesis (Loftus et al., 1985) from the perspective of memory reconsolidation theory (J. C. K. Chan & LaPaglia, 2013; Hardt et al., 2010). According to this account, reactivating event memory traces will trigger a time-dependent reconsolidation process during which they are vulnerable to modification through the presentation of misinformation. However, decades of intensive investigation has revealed that the misinformation effect is strongly moderated by the precise nature of the test used to measure participants' knowledge of the event (Loftus, 2005; Zaragoza et al., 2006). A previous study that reported reconsolidation-mediated overwriting in a misinformation paradigm (J. C. K. Chan & LaPaglia, 2013) arguably employed a test (the True/False Test) unsuitable for the detection of trace alteration because it does not exclude the contribution of factors unrelated to memory impairment (Belli, 1989; Lindsay & Johnson, 1989; Zaragoza & McCloskey, 1989).

In three reconsolidation-misinformation experiments, we employed an alternative assay of participants' event knowledge (the Modified Test; M. McCloskey & Zaragoza, 1985a), which is considered to minimize the role of such factors, thus providing a relatively 'pure' index of memory impairment (either storage- or retrieval-based; see Figure 4.1). Classic misinformation effects were replicated with conventional tests (Original Test, True/False Test), but were not observed with the Modified Test. This was also the case when different reminder protocols were employed (Experiment 10, cf. Hupbach et al., 2008; Forcato et al., 2009). The conclusion of 'no impairment' in previous Modified Test studies has always been somewhat tentative (Zaragoza et al., 1992) because it is inherently difficult to interpret null effects within the NHST framework (Dienes, 2014). However, we overcame this ambiguity by using Bayes Factors to determine the degree of evidentiary support for the null hypothesis relative to the reconsolidation-overwriting hypothesis. Together these experiments provide compelling evidence against the claim that reconsolidation enables the overwriting of event memories in the misinformation paradigm (J. C. K. Chan & LaPaglia, 2013; Hardt et al., 2010).

The findings observed here do not suggest that retrieving information prior to new learning has no influence on participants' memory, only that it does not enable the overwriting of event memory traces. In fact, there is considerable evidence that retrieval practice can both enhance subsequent recall of previously acquired 'old' information (the "backward testing effect"; Tulving, 1967; Roediger & Butler, 2011) and simultaneously enhance learning of 'new' information (the "forward testing effect"; Pastötter & Bäuml, 2014). The latter phenomenon has been investigated specifically in the context of the misinformation paradigm where it is has been referred to as 'retrieval enhanced suggestibility' (RES; L. T. Gordon, Thomas, & Bulevich, 2015; J. C. K. Chan et al., 2009; J. C. K. Chan & LaPaglia, 2011; J. C. K. Chan & Langley, 2011). A RES effect occurs when, relative to a group receiving no reminder, a retrieval practice phase prior to misinformation exposure increases the likelihood that participants' will respond with the misleading information on a subsequent test.

It was not our intention to evaluate the RES hypothesis in the present experiments; indeed RES effects can only be examined with tests that allow for responding with the misled item. RES effects were therefore not anticipated on the Modified Test and True/False Test, which were the focus of the present investigation[37]. Nevertheless, evidence obtained in recent RES studies is consistent with the findings of the present experiments. L. T. Gordon and Thomas (2014) measured event knowledge using the Modified-Modified-Free-Recall test (MMFR; J. M. Barnes & Underwood, 1959) which allows participants to report details from both the event and post-event stages. Compared to a no-reminder group, participants who performed a retrieval practice task prior to misinformation exposure recalled more misled items *and* event items. Therefore, prior retrieval of event information appears to enhance the learning of subsequently presented misinformation, but not at the expense of participants' ability to recall the event information, a finding entirely consistent with the persistence hypothesis.

Outside the context of memory reconsolidation, is there remaining evidence for the overwriting hypothesis? In the introduction we referred to a previous meta-analysis of misinformation studies that employed the Modified Test. Payne et al. (1994) found that although in the majority of experiments (30/44) researchers did not observe misinformation effects, when all cases were combined, there was a small but statistically significant difference between the control (M = 75.8%) and misled (M = 71.7%) conditions, $t(43) = 4.65$, $p < .001$. If accurate, this analysis suggests that memory impairment, and thus potentially trace overwriting (see Figure 4.1), does contribute to misinformation effects, albeit only to a small degree. However, the analysis approach is potentially problematic. Treating each experiment as an individual data point and conducting a single overall t-test does not account for the precision of each

[37]A no-reminder control group would be required to measure RES using the Original Test in Experiment 8.

effect size estimate. Formal meta-analytic methods typically attribute greater weight to studies with higher precision, as they more accurately estimate the magnitude of the effect under scrutiny (Cumming, 2012). It is possible that Payne et al. (1994) were unable to account for precision in their analysis because standard deviations were not reported in the majority of the original sources. Indeed, this issue has also prevented us from running our own meta-analysis using models with appropriate precision-based weighting. However, as we were able to obtain information about sample size from the original sources, it was possible to make some assessment of the degree to which precision could be influencing the magnitude of misinformation effects in this particular set of studies.

Figure 4.4 shows a scatter plot of sample size against the magnitude of the misinformation effect for Experiments 8-10 and the 41 experiments[38] used in the analysis by Payne et al. (1994). Visual inspection suggests an asymmetrical distribution: moderate misinformation effects have been detected in experiments with smaller sample sizes, but are absent from experiments with particularly large sample sizes. A relationship between study precision and effect magnitude is often quantified using Egger's test (Egger, Davey Smith, Schneider, & Minder, 1997) to evaluate whether the pattern of data points is indicative of selective reporting bias. The test computes a regression of the standardized effect size against inverse standard error (as a proxy for study precision), weighted by the inverse of the variance. However, because we lack the necessary statistics (*SD*s, *SE*s, or *t*s) to perform this test, we have used sample size as a proxy for study precision.

A regression model (weighted by sample size; see Figure 4.4) indicates that the magnitude of the misinformation effect has a modest relationship with sample size ($R^2 = 0.12$, $t =$ -2.41, $p = 0.02$). The reasons for this pattern are difficult to identify definitively, and do not necessarily reflect selective reporting bias (Sterne & Egger, 2001). Distribution asymmetry can also be driven by other characteristics of experiments that happen to have small samples. Figure 4.4 does appear to indicate that larger misinformation effects (>.1) have typically only

[38] Three cases used in the original analysis (Payne et al., 1994) were subgroups within experiments (Belli, 1993; Bonto & Payne, 1991). As sample size was reported at the experiment level in the original sources, we have collapsed over these subgroups resulting in 41 rather than 44 cases overall.

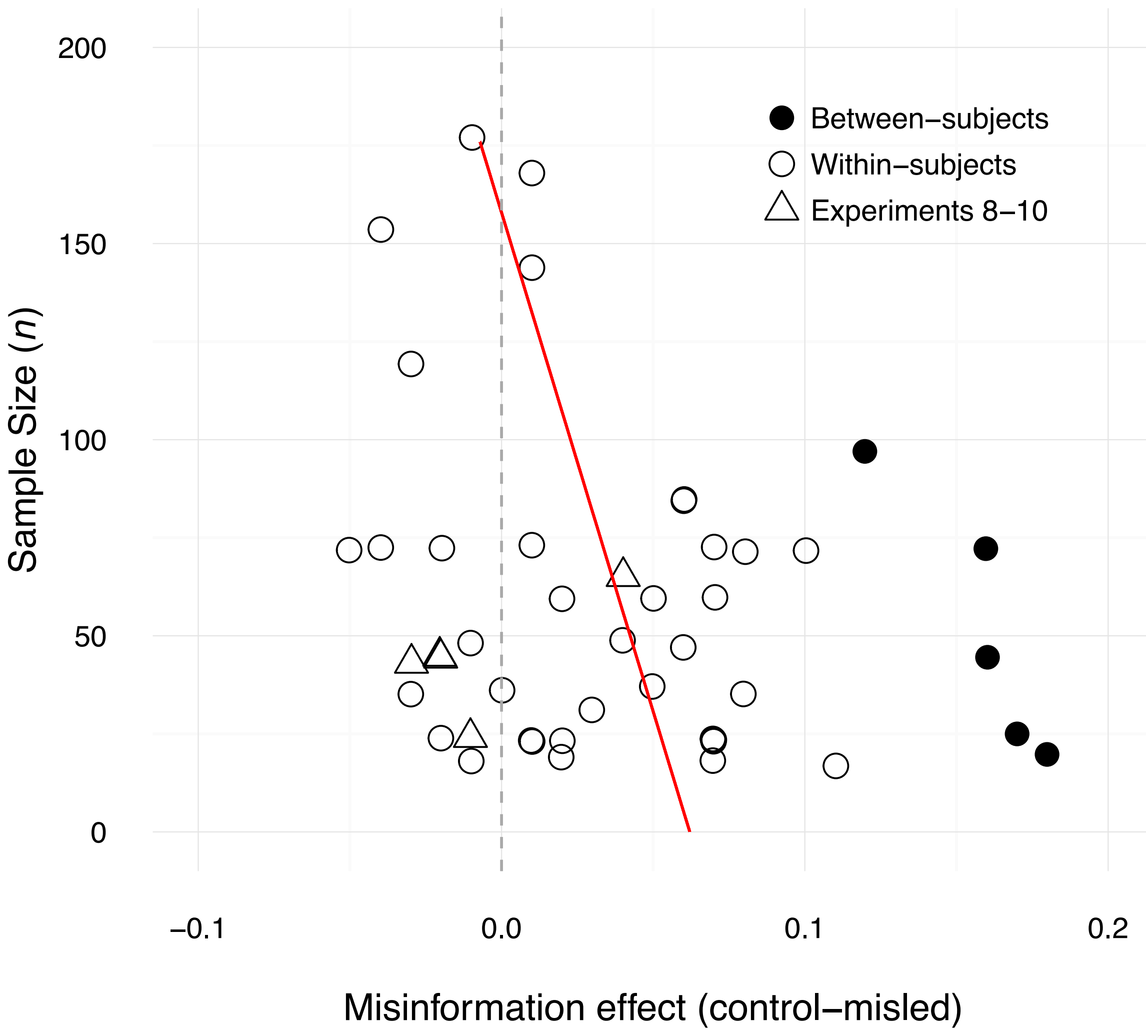


Figure 4.4: Scatterplot showing the 41 Modified Test experiments (circles) used in the meta-analysis by Payne et al. (1994) and Experiments 8-10 (triangles) as a function of sample size (n) and misinformation effect (control performance minus misled performance). Black shapes indicate between-subject designs and white shapes indicate within-subject designs. The red line represents a weighted regression model (see main text for details). The dashed grey line represents a zero effect.

been observed in between-subjects designs which are likely to be underpowered compared to their within-subject counterparts (assuming equivalent sample size). It is worth noting that all 5 of these between-subjects experiments were reported in two infant studies (Ceci, Ross, & Toglia, 1987; Toglia, Ross, Ceci, & Hembrooke, 1992), and similar effects were not observed in close replication attempts that employed a within-subjects design (Zaragoza et al., 1992). Additionally, 7 of the experiments that observed modest and statistically significant effects were drawn from a line of empirical enquiry that ultimately found evidence that the performance deficits were temporary, and therefore could not be due to trace overwriting (Chandler, 1989, 1991, see below for further details).

In light of strong evidence supporting the null hypothesis (e.g., Experiments 8-10, M. McCloskey & Zaragoza, 1985a), we have naturally sought to identify reasons why misinformation effects were observed in a small number of Modified Test studies. However, the reader is cautioned that our observations and analyses pertaining to the previous meta-analysis (Payne et al., 1994) are necessarily exploratory (i.e., analyses were conducted *after* looking at the data), and therefore do not constitute a confirmatory test of a pre-specified hypothesis (Wagenmakers et al., 2012). Nevertheless, the apparent relationship between effect magnitude and study precision precludes a straightforward interpretation of any overall effect size estimate one may attempt to derive from this data set, particularly if that estimate does not account for study precision as was the case in Payne et al. (1994). Unfortunately, the absence of statistical information in the original sources has proved a major obstacle, and it remains an open question as to whether impairment effects detected using the Modified Test are reliable. In order to enable future endeavours towards cumulative science, we would strongly encourage researchers to make their raw data publicly available in a persistent third party repository upon publication (Morey et al., 2016). Going forward, the field might consider conducting a pre-registered, multi-lab, Registered Replication Report (Simons et al., 2014) to identify the degree to which memory impairment effects can be reliably detected in the misinformation paradigm.

It is worth repeating that even if misinformation effects are observed using the Modified Test, such a finding would only be consistent with the overwriting hypothesis but cannot determine the relative involvement of storage-based and/or retrieval-based factors (see Figure 4.1). Furthermore, there is an unfortunate philosophical asymmetry between storage-based and retrieval-based accounts. Given the limited methodological repertoire of the experimental psychologist (M. J. Watkins, 1990), and our incomplete theoretical characterisation of the memory trace (for a recent review see Josselyn, Köhler, & Frankland, 2015), storage-based accounts cannot be directly corroborated at present. As such, only retrieval-based accounts currently provide testable predictions, and proponents of storage-based accounts are left in the unenviable position of claiming theory corroboration only when retrieval-based explanations appear to have been exhausted (e.g., Melton & Irwin, 1940).

Unfortunately, the role of retrieval-based factors is not always immediately apparent. The research line developed by Chandler[39] provides a clear illustration of this issue (Chandler, 1991, also see 1993; and Windschitl, 1996). After observing retroactive interference effects using the Modified Test, Chandler (1989) initially concluded that her findings offered support for the overwriting hypothesis. However, in subsequent experiments[40], Chandler (1991) observed that the interference did not persist over a 48hr delay, a 'spontaneous recovery' effect (Rescorla, 2004) that can presumably be attributed to the alleviation of some transient inhibitory process, and demonstrates that overwriting had not occurred. It is striking that across a variety of retroactive interference paradigms, performance deficits observed on an initial test can often be alleviated, after waiting for a short delay, or through the provision of additional retrieval cues (Bouton, 2002; Tulving, 1974, see Chapter 1).

The need to consider retrieval-failure explanations is especially pertinent for reconsolidation

[39]Chandler's stimuli were deliberately designed to be somewhat different to those typically used in misinformation studies (for details see Chandler, 1989). The 'event', 'misled', and 'foil' items were adjacent sections of landscape scenes (e.g., a beach) divided into three parts. Because the scenes were fairly homogeneous, the subsections were highly similar and not easily assigned distinctive verbal labels (unlike for instance, "stop" sign and "yield" sign in Loftus et al., 1978). Chandler (1989) argued that this reduced discriminability was more likely to elicit interference effects. In addition, unlike many misinformation studies both the event information and post-event information were presented in the same modality, likely making them less discriminable during the Final Test Phase (also see Lindsay, 1990).

[40]Chandler (1991) also reported several replications from her laboratory, including a within-experiment comparison of the short- and long-retention interval groups.

theory, which suggests that under specific conditions 'recovery' from retroactive interference can be avoided (Nader, 2003a; Nader & Hardt, 2009), a claim that has considerable ethical and clinical ramifications (Hui & Fisher, 2014; Schwabe et al., 2014). Moreover, our review of the literature indicates that the extant evidence for the theory is remarkably tenuous (see Chapter 2). Although striking performance deficits have been reported in human participants using a procedural sequence learning paradigm (Walker et al., 2003), and a fear-conditioning paradigm (Schiller et al., 2010, 2013), these effects were absent in several subsequent replication attempts (e.g., Chapter 3, Hardwicke et al., 2016; Golkar et al., 2012; Kindt & Soeter, 2013), suggesting that either the original effects were false positives, or the necessary conditions for reliably obtaining reconsolidation effects have not yet been adequately specified. Either way, future replications are to be encouraged as they are a vital tool for verifying the reliability of previous findings, and/or identifying genuine boundary conditions, both essential ingredients of theoretical progress. In Experiment 9, we obtained an approximate replication of the reconsolidation effect reported in Chan & LaPaglia (2013), but found that it is contingent on the type of test used to assay participants' event knowledge, and therefore cannot be attributed to reconsolidation.

In summary, the findings of the present study have demonstrated that although exposure to contradictory misinformation can influence performance on some types of memory test, this cannot necessarily be attributed to reconsolidation-mediated overwriting of existing memory traces. This does not diminish the potential implications of the misinformation effect for eyewitness testimony. In fact, our findings imply a degree of optimism: if event memory traces have persisted then the information that they represent has not been irrecoverably lost. This highlights the importance of fostering a favourable retrieval environment (Chandler & Fisher, 1996), and critically, asking the witness the right question.

Chapter 5

Does memory retrieval help or hinder the correction of semantic misconceptions?

> "Some beliefs...we shall probably retain while we live. Some...we may abandon tomorrow in the face of adverse evidence."
>
> — Quine and Ullian, 1978

Often we are quite adept at answering general knowledge questions like "Who wrote the novel '1984'?" (Orwell), "What is the capital of Peru?" (Lima), and "What is the collective term used to describe a group of owls?" (a parliament). However, it is not uncommon for errors to infiltrate our knowledge store. Individuals readily learn erroneous information from a diverse range of sources, including films (Butler, Zaromb, Lyle, & Roediger, 2009), fictional stories (Marsh, Meade, & Roediger, 2003), and even multiple-choice tests (Roediger & Marsh, 2005). In one survey of 671 university students for example (Tauber, Dunlosky, Rawson, Rhodes, & Sitzman, 2013), it was relatively common for individuals to suggest that the capital of Australia is Sydney (it's Canberra), the largest desert on earth is the Sahara (it's Antarctica), and the name of the man who started the Reformation in Germany was Hitler (it was Luther).

Although misconceptions can be fairly innocuous, they can also have serious detrimental consequences at both the individual and societal level (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012b). For example, in 1998 a (subsequently retracted) study published in *The Lancet* claimed to have identified a link between the measles-mumps-rubella vaccine and autism. The subsequent dissemination of this misinformation was implicated in reduced vaccination rates, and outbreaks of previously controlled diseases, a situation described as a 'public health tragedy' (Gust, Darling, Kennedy, & Schwartz, 2008; Poland & Spier, 2010). In education, several surveys have revealed a high prevalence of brain-related misconceptions ('neuromyths') amongst teachers, for example, "Short bouts of co-ordination exercises can improve integration of left and right hemispheric brain function" (Dekker, Lee, Howard-

Jones, & Jolles, 2012). This false information is often used to justify the use of discredited educational interventions in classrooms (Howard-Jones, 2014). Despite concerns about the harmful effect of misconceptions, there remains a great deal to learn about methods for effectively 'correcting' or 'updating' erroneous 'old' information stored in memory with replacement 'new' information.

An improved understanding might be gained through consideration of the widely reported empirical phenomenon of 'retroactive interference', whereby learning of new information perturbs the subsequent recall of previously learned old information (M. C. Anderson & Neely, 1996; Crowder, 1976). This effect has been intensively investigated using the A-B, A-C paired-associate learning paradigm (for details see Chapter 1) whereby cue-target pairs (e.g., *Horse-Leaf*; A-B) are 'updated' with new targets (e.g., *Horse-Chair*, A-C). This is broadly analogous to replacing a semantic misconception (e.g., *Capital of Australia-Sydney*, A-B), with accurate information (e.g., *Capital of Australia-Canberra*, A-C)[41]. On a subsequent test in which participants are provided with cues and asked to respond with targets, it is typically found that recall of the old information (B) is markedly impaired relative to a control group who did not learn the new information (C; e.g., Underwood, 1948; Melton & Irwin, 1940). According to this evidence then, simply providing replacement information might reduce the impact of semantic misconceptions by impeding their subsequent recall.

A fundamental issue however, is that retroactive interference effects are often temporary and depend on the precise nature of the test being used to assay memory (also see Chapter 4). With the passage of time, retroactive interference in the paired-associates paradigm will typically diminish, and old responses can recover their dominance (Briggs, 1954; A. S. Brown, 1976; Wheeler, 1995). In addition, asking participants to provide *both* responses (B and C; J. M. Barnes & Underwood, 1959), or recognise the old information (B) amongst distractors (e.g., D, E, F; R. C. Anderson & Watts, 1971; Postman & Stark, 1969), can attenuate or even eliminate the effect, demonstrating that the memory traces representing the old information

[41]The situation will likely be substantially more complicated for semantic knowledge that is embedded within a larger associative network and/or forms an integral part of one's belief system/worldview (see Lewandowsky et al., 2012b for review).

are often intact (M. C. Anderson & Neely, 1996; Crowder, 1976). Similar outcomes have been reported in paradigms involving somewhat more complex stimuli. For example, the classic 'misinformation effect' discussed in Chapter 4 can be eliminated through the use of an appropriately designed recognition or recall test (M. McCloskey & Zaragoza, 1985a; Zaragoza, McCloskey, & Jamis, 1987). Therefore, retroactive interference effects appear to reflect a temporary shift in the relative dominance of different cue-target associations, and do not necessarily entail the permanent replacement of the old information. From a normative perspective such an outcome might be deemed sub-optimal: an updating process should ideally *erase* the erroneous information such that it can no longer continue to influence an individual's behaviour (Bjork & Bjork, 1996; Lewandowsky et al., 2012b; Seifert, 2002).

Recently, theorists have argued that the veracity and relevance of information stored in memory is routinely maintained through the brain's inherent capacity for neural plasticity (Dudai, 2009; J. L. Lee, 2009). According to these accounts, the act of retrieval destabilises existing memory traces, triggering a time-dependent molecular restabilisation process known as 'reconsolidation' (for details see Chapters 1, 2, and 3; Nader & Hardt, 2009). During this temporary period of vulnerability, there is an opportunity for existing memory traces to be 'updated', 'modified', 'erased', or 'overwritten' when learning new information disrupts the reconsolidation process (J. C. K. Chan & LaPaglia, 2013; Dudai, 2009; J. L. Lee, 2009; Nader, 2003a; Schiller & Phelps, 2011). Therefore, reconsolidation theory suggests that semantic misconceptions could be permanently impaired, or even eliminated, by retrieving them immediately prior to learning replacement information.

Although reconsolidation theory implies that memory retrieval should help with the correction of semantic misconceptions, a considerable body of evidence suggests that retrieval could actually be more of a hindrance. The 'backward testing effect[42]' (Carrier & Pashler, 1992; Payne, 1987; Roediger & Butler, 2011) refers to the phenomenon whereby actively retrieving information ('retrieval practice') from memory increases the likelihood of its successful recall

[42]The phenomenon is typically referred to simply as 'the testing effect', however we use 'backward testing effect' (i.e., benefits of retrieval practice for previously learned old information) here to distinguish it from the 'forward testing effect' (i.e., benefits of retrieval practice for to-be-learned new information) described in the following paragraph.

on a subsequent test, relative to simply re-studying the same information, which in turn is often (but not always) more beneficial than not reviewing the material at all. Contrary to the predictions of reconsolidation theory, this evidence raises the possibility that retrieving semantic misconceptions will actually reinforce them, rather than eliminate them.

Intriguingly, several studies indicate that prior retrieval of 'old' information can also enhance the learning of subsequently presented 'new' information. Within the framework of the misinformation paradigm, such effects are termed 'retrieval-enhanced suggestibility' (see Chapter 4; J. C. K. Chan et al., 2009), but the more general phenomenon has been referred to as the 'forward testing effect' (Pastötter & Bäuml, 2014; also see Tulving & Watkins, 1974). A recent study using a misinformation paradigm and a testing procedure that assessed recall of both old and new items found that forward and backward testing effects are not necessarily in conflict: retrieval practice can simultaneously enhance the subsequent recall of previously learned old information *and* to-be-learned new information (L. T. Gordon & Thomas, 2014).

In summary, the forward and backward testing effects (from herein 'hybrid testing account') suggest that reviewing old information, especially through retrieval practice, has the potential to both help and hinder the correction of semantic misconceptions: retrieval may enhance learning of new replacement facts, but it could also reinforce the old erroneous information. The aforementioned reconsolidation theory also suggests that retrieval will facilitate the correction process, however it proposes that knowledge updating will be realised through the overwriting of the memory traces representing the old information. Thus, the two accounts make competing predictions about the persistence of semantic misconceptions after they have been replaced with new information.

One previous study that directly addressed this conflict is especially pertinent to the current investigation (Pashler et al., 2013). Critically, this study followed the canonical distributed timing protocols necessary for a controlled investigation of reconsolidation theory, separating each critical stage (training, retrieval/new learning, final test) by at least a day (see Chapter 2). In Experiment$_P$ 1 ($n = 56$), participants were first taught 36 new 'facts' about several fictitious topics (Day 1). For example, participants learned that the diet of the 'Golden-Eared Marmoset' mainly consists of beetles. On Day 2, participants first re-studied 12 facts, were tested ('retrieval practice') on 12 other facts, and did not review the remaining 12 facts. Subsequently, they were told that all of the previous information they had been taught was incorrect, and presented with replacement information for each of the facts. For example, in the case of the Golden-Eared Marmoset, participants were told that actually their diet consists mainly of tree sap. Finally, on Day 8, participants were asked to recall the *replacement* information for each fact. They were explicitly instructed not to recall the Day 1 information.

The main finding was that accuracy on the Day 8 test (i.e., responding with the replacement fact) was higher in the two reminder conditions (retrieval practice, restudy) compared to the no-reminder condition (rest). In other words, consistent with both the hybrid testing account and reconsolidation theory, updating of erroneous semantic knowledge was facilitated by the provision of a reminder prior to learning the replacement information (although retrieval practice did not appear to confer any additional advantage over restudy). Because participants were explicitly discouraged from responding with the old facts, this experiment could not ascertain the fate of the knowledge that they had apparently abandoned. According to the hybrid testing account, this information should have been enhanced whilst according to reconsolidation theory it should have been eliminated. Pashler et al. addressed this issue in a second experiment (Experiment$_{\mathrm{P}}$ 2, n = 69) where participants followed a similar procedure to Experiment$_{\mathrm{P}}$ 1, but were now asked to recall facts from *both* Day 1 and Day 2 on the Day 8 test. The findings replicated and extended those observed in Experiment$_{\mathrm{P}}$ 1. Recall of both Day 1 and Day 2 facts was higher in the two reminder conditions (retrieval practice, restudy) compared to the no-reminder condition (rest).

These findings appear to demonstrate that retrieval can facilitate learning of new replacement facts and simultaneously reinforce previously learned old facts (also see L. T. Gordon & Thomas, 2014). From the perspective of the hybrid testing account, it is surprising that there was no additional enhancement gained through retrieval practice relative to restudy in either experiment. However, this could be due to the use of a blocked design in which all items were reviewed together (restudy and retrieval practice; block 1) before corrective feedback was provided (block 2). This would disrupt any close connection between the review procedure used for each old fact and the subsequent learning of a new fact. Nevertheless, the findings are clearly contrary to the reconsolidation-overwriting hypothesis, which predicted an effect in the opposite direction. In the present study, we sought to expand upon these findings by attempting to correct genuine semantic misconceptions (rather than recently learned fictitious facts) that individual participants had learned naturally in their day-to-day lives (see Figure 5.1).

On Day 1, participants completed 120 general knowledge questions and their responses were classified as “correct”, “incorrect”, or “skipped”. All of the participants’ incorrect responses, and a subset of their correct responses, were carried forward to Day 2. During this stage, participants cycled through the questions, and completed a reminder phase followed by a corrective feedback phase for each item in turn (i.e., an interleaved rather than a blocked design, cf. Pashler et al., 2013). The critical manipulation came during the reminder phase: for some items, participants were asked to recall their previous Day 1 response (‘retrieval practice’); for other items, they were asked to restudy their previous Day 1 response; and for the remaining items, they simply rested. On Day 3, participants cycled through each of the questions again (as on Day 2, these were only the items answered incorrectly on Day 1) twice in two separate phases (blocked): firstly, in the ‘dominance phase’ they were asked to respond to the questions with whatever answer they thought to be correct. Secondly, in the ‘persistence phase’, they were asked to specifically provide both their previous Day 1 response, and the replacement fact they had learned on Day 2 (as in Experiment$_{\mathrm{P}}$ 2). Thus for each item, we aimed to evaluate the extent to which there was a shift in dominance from ‘old’ to ‘new’ responses (i.e., correction/updating of semantic misconceptions), the extent to which this entailed destructive loss or persistence of the old responses, and the extent to which responses in both test phases were influenced by the reminder manipulation. Finally, in order to examine potential ‘spontaneous recovery’ effects (Wheeler, 1995), we repeated the Day 3 testing regime on a delayed test occurring approximately 3 weeks later.

Based on both reconsolidation theory and the hybrid testing account, we expected that memory retrieval would facilitate the updating of semantic misconceptions. Specifically, during the dominance phase, as well as the persistence (new) phase, we expected higher performance (responding with replacement facts) in the reminder conditions compared to the no-reminder condition (retrieval practice $\geq$ restudy $>$ rest). Note that the hybrid testing account also implies higher performance in the retrieval practice vs restudy conditions, but reconsolidation does not necessarily predict a difference between these conditions as both will presumably involve reactivation of target memory traces. The critical test on which the two accounts offer contrasting predictions is the persistence (old) phase. Reconsolidation theory predicts a retrieval-induced performance *decrement* (rest $>$ retrieval practice $\geq$ restudy; H_1) whereas the hybrid testing account anticipates a retrieval-induced performance *increment* (rest $<$ restudy $<$ retrieval practice; H_2).

5.1 Experiment 11

5.1.1 Methods

All experimental materials, data, and analysis code pertaining to this chapter have been made publicly available on the Open Science Framework (https://osf.io/njywe). All data exclusions, manipulations, and measures (aside from confidence ratings) conducted during this study are reported.

5.1.1.1 Participants

Fifty-nine participants (29 female; 30 male; age *Mdn* = 27 years, range = 18-39 years) were recruited from the online crowd-sourcing platform Prolific Academic, and took part in the experiment remotely using their own personal computing devices. All participants reported that they spoke fluent English and met the other inclusion criteria outlined below. Participants provided informed consent and received monetary compensation. The local UCL ethics committee approved the study.

To ensure basic levels of data quality, we pre-specified a number of inclusion criteria, and programmed our software to automatically exclude individuals who failed to comply. The criteria were as follows (with n additional individuals who did not meet the criteria in brackets): navigating to other web pages or software more than three times during a stage (n = 25), starting but not completing at least stages 1-3 (n = 15), reporting during debrief that they had looked up question answers ($n = 6$), failing to complete a stage within 2 hours ($n = 1$), failing to provide a minimum of 9 incorrect responses during Stage 1 ($n = 1$), responding incorrectly to two or more 'catch' trails per stage ($n = 0$), or failing to complete each stage within 24hrs (+/- 3 hrs) of the previous stage ($n = 0$). Participants were reminded of these criteria at the start of each stage (see Appendix H).

5.1.1.2 Materials

120 general knowledge questions and their correct answers were used as stimuli (see Appendix G). Some of these were selected because they had elicited a relatively high number of commission errors in a previous norming study (Tauber et al., 2013), for example, "What is the name of the largest desert of earth?", (correct answer: 'Antarctica', common commission error: 'Sahara'). We also included several relatively straightforward questions in order to keep participants motivated, for example, "What is the name of the horse-like animal with black and white stripes?", (correct answer: 'zebra'). Additional questions were drawn from various Internet sources.

5.1.1.3 Design

The experiment was a 3 (reminder type: rest, restudy, retrieval practice) x 2 (final test time: Day 3, delayed) repeated-measures factorial design (see Figure 5.1). In the dominance phase of testing, participants were asked to provide the correct response (i.e., the replacement fact) and in the persistence phase they were specifically asked to provide their old response (i.e., their Day 1 response), and the replacement fact from Day 2. Therefore, the three primary dependent variables were the proportion of correct responses during the dominance phase and persistence phase (old and new). We also examined the proportion of intrusions, specifically, old facts provided in the dominance phase, new facts provided for the old probe in the persistence phase, and old facts provided for the new probe in the persistence phase.

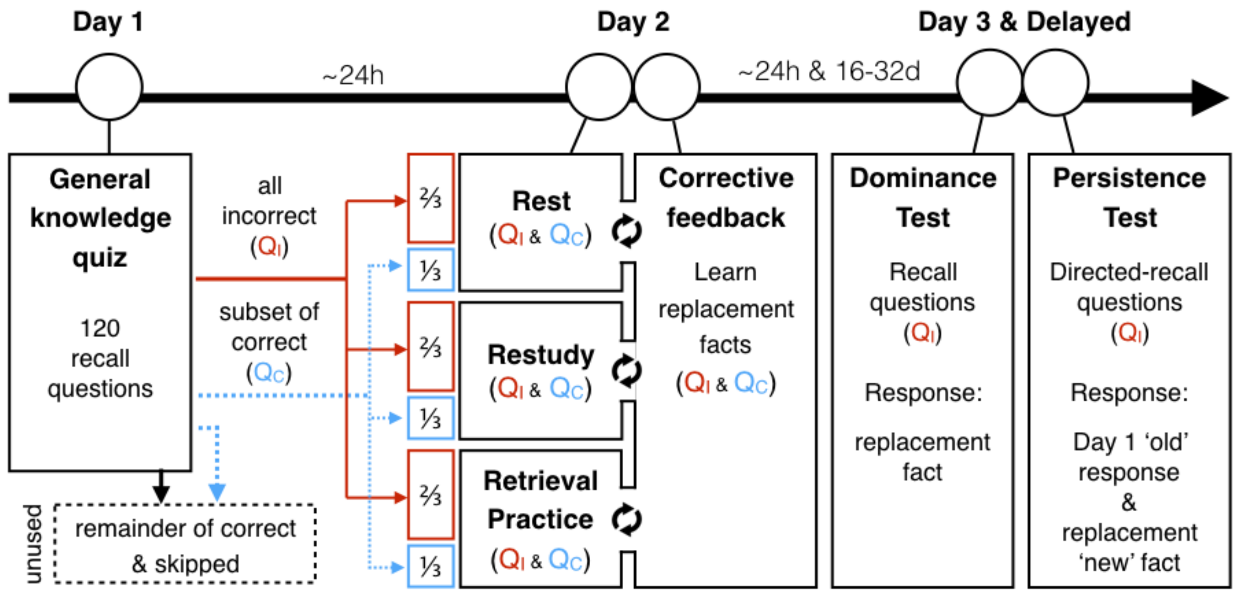


Figure 5.1: Study design depicting the stages of the experiment in temporal order from left to right. Red indicates usage of questions that were responded to incorrectly during Day 1 (Q_I; i.e., semantic misconceptions). Blue indicates usage of a random subset of questions that were responded to correctly during Day 1 (Q_C). See main text for details.

5.1.1.4 Procedure

Participants began by reading general information about the study (Appendix H) and were informed that "We are investigating how people learn and remember information. In this particular study we will be asking you general knowledge questions and providing you with correcting information if you make errors."

On Day 1 (see Figure 5.1), participants responded to 120 sequentially presented general knowledge questions by typing their answers into a text box. No feedback was provided and responses were untimed. Participants could skip to the next question if they did not know the answer.

At the end of this stage, each question was classified by a computer algorithm as having been skipped, receiving a correct response, or receiving an incorrect response. This was achieved using a 'string-matching' algorithm that computed an index of similarity (0-1) between the correct answer and the response entered by the user[43]. Responses with a similarity score greater than or equal to .8 were considered correct as this typically reflected that the response contained some minor spelling errors but was broadly accurate.

For each participant, all of the questions that they responded to incorrectly (i.e., similarity score < .8) were used on Day 2 and Day 3 (see Figure 5.1). In order to make the task less onerous, these were supplemented with a randomly selected subset of the questions that had been responded to correctly. Consequently, two-thirds of the questions used during Days 2 and 3 had been responded to incorrectly during Day 1 and the remaining third had been responded to correctly. This ratio was approximately maintained after the questions were randomly allocated to the three reminder conditions (see Table 5.1).

During Day 2, each trial consisted of a reminder phase (20s) followed by a corrective feedback phase (20s). The reminder phase involved participants either being instructed to rest (Rest condition), re-study the question and response they gave during Day 1 (Restudy condition), or provide the response they gave during Day 1 (Retrieval Practice condition). During the corrective feedback phase, participants were informed whether their Day 1 response was correct or incorrect, and in the latter case, provided with the correct replacement fact.

On Day 3, there were two blocked phases of testing: (1) a ‘dominance’ phase intended to assess whether participants had ‘updated’ their knowledge by replacing their previous erroneous response with the replacement fact as their preferred response to each question; and (2) a ‘persistence’ phase intended to assess whether both the erroneous and replacement information was retrievable from the memory store. During the ‘dominance’ phase, the same questions as used on Day 2 (i.e., items responded to incorrectly on Day1) were presented in the same format employed during Day 1. After all of the questions had been responded to, the ‘persistence’ phase began, during which participants were asked to provide two responses to each question: the original ‘old’ response they gave on Day 1, and the ‘new’ replacement fact

[43] Specifically, this was a normalized ratio of the Levenshtein edit distance, a metric which identifies the number of “fundamental” operations (substitution, deletion, insertion, or transposition) needed to convert one character string into another, thus providing an index of their “similarity”. The computer code is available on the Open Science Framework (https://osf.io/etd5m/).

they were given on Day 2. All responses were untimed. Approximately 3 weeks (median = 24 days; range = 16-32 days) after they had completed the Day 3 test, a subset of participants ($n = 44$) returned to take part in another ‘delayed test’. This session employed the same procedure as Day 3.

During each stage, questions were presented in a random order. It was not possible to return to a question once it had been skipped or a response had been submitted. Each stage also included three ‘catch trials’ appearing at approximately one-third intervals. Participants were simply asked to type in a 5-item code that was presented on the screen (e.g., 3RYX7). These trials were included to deter participants from rapidly clicking the skip/submit button in order to cycle through questions as quickly as possible without paying attention (see Oppenheimer, Meyvis, & Davidenko, 2009).

5.1.2 Results

5.1.2.1 Day 1

Metrics for individual questions (Qs) are available in Figure 5.2 (Qs 1-60) and Figure 5.3 (Qs 61 -120; question numbers can be matched to verbatim questions in Appendix G). Overall, participants responded to almost half of the general knowledge questions correctly (M = 0.46, SD = 0.13; blue bars), almost a quarter incorrectly (M = 0.23, SD = 0.09; red bars), and skipped around a third (M = 0.33, SD = 0.14; yellow bars).

As intended, the majority of participants found some questions relatively easy (e.g., “What is the capital of France?”, Q24), whereas other questions elicited a large number of commission errors (e.g., “What was the last name of the first European explorer to land in North America?”, Q105). Some questions (e.g., “What is the name of the small Japanese stove used for outdoor cooking?”) had high skip rates and may be worth excluding from future experiments.

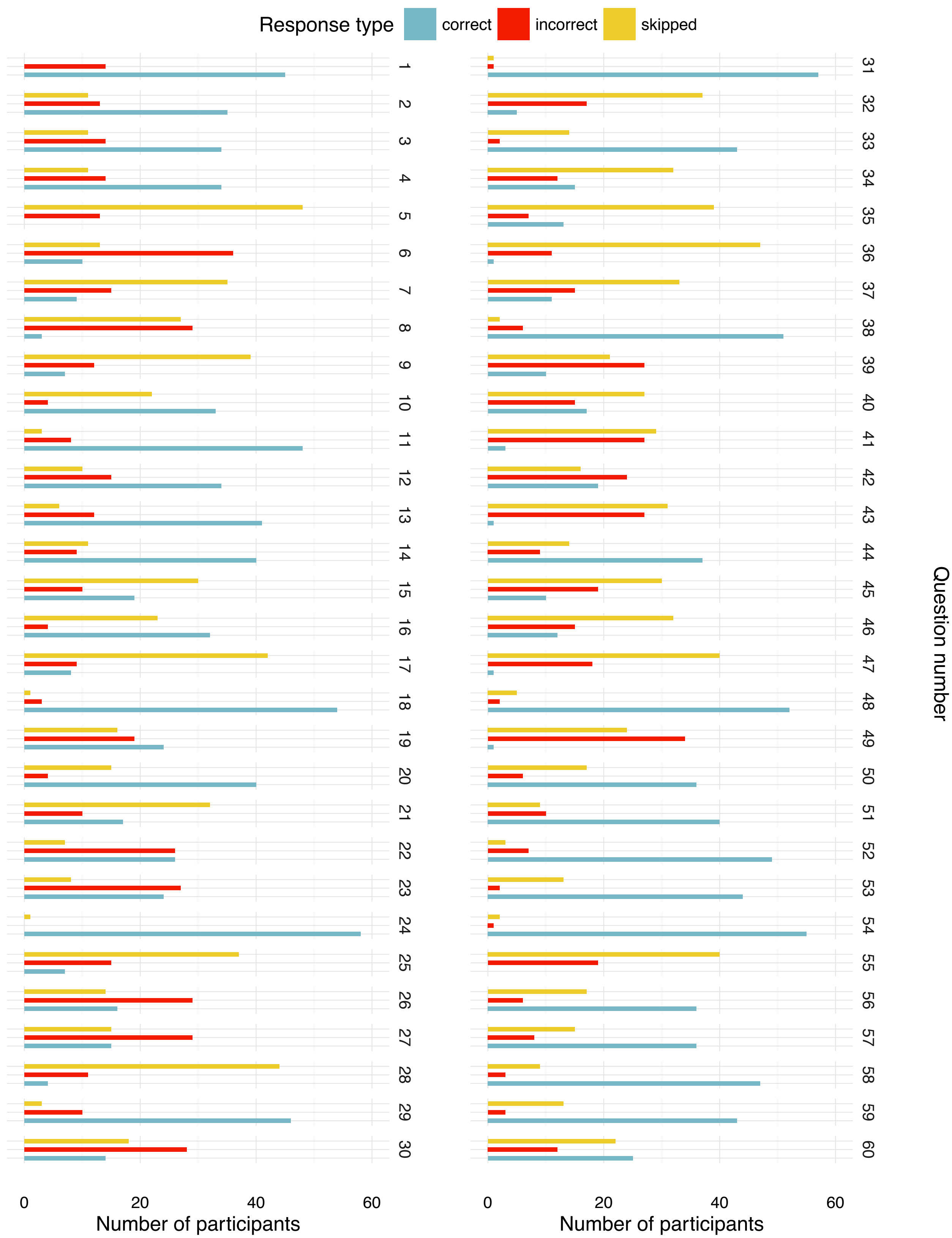


Figure 5.2: Number of correct, incorrect, and skipped responses for questions 1-60 on the Day 1 test. Verbatim questions can be matched to their numbers in Appendix G.

Figure 5.3: Number of correct, incorrect, and skipped responses for questions 61-120 on the Day 1 test. Verbatim questions can be matched to their numbers in Appendix G.

Table 5.1: Item allocation across the three reminder conditions (rest, restudy, retrieval practice) as a function of Day 1 response type (correct, incorrect).

Reminder condition	Day 1 response type	Items allocated $M(SD)$
rest	correct	5.15 (1.91)
rest	incorrect	9.66 (3.77)
restudy	correct	4.49 (1.78)
restudy	incorrect	8.95 (3.72)
test	correct	4.71 (1.80)
test	incorrect	9.37 (3.70)

5.1.2.2 Day 2

For each participant, incorrect items and a subset of correct items from Day 1 (see procedure) were carried forward to Day 2. Items (questions) were randomly assigned to the three reminder conditions (rest, restudy, retrieval practice). Because the number of incorrect items for a given participant was not necessarily divisible by 3, it was inevitable that the reminder conditions were not strictly balanced. However, as shown in Table 5.1, the number of items allocated to each condition was approximately equal.

During Day 2, participants provided responses in the retrieval practice condition which afforded an opportunity to assess the degree to which responding was stable between Day 1 and Day 2. We used the string-matching algorithm (see Procedure) to assess the degree of similarity between Day 2 and Day 1 responses. Responses were classified as a 'match' if the similarity score was greater than or equal to .8, and 'mismatch' otherwise. The majority of responses were matches (proportion 'match' $M = 0.89$, $SD = 0.13$) demonstrating substantial stability between tests.

5.1.2.3 Day 3 Test and Delayed Test

For the Day 3 Test and Delayed Test we were interested in the proportion of correct responses (‘accuracy’), the proportion of intrusions (‘intrusions’), and whether these variables differed as a function of Day 2 reminder type (rest, restudy, retrieval practice) and final test time (Day 3, Delayed; see ‘design’ section). We began by running a series of 3x2 repeated measures ANOVAs with reminder type (rest, restudy, retrieval practice) and test time (Day 3, delayed) as independent variables, and either accuracy or intrusions as dependent variables. These analyses were restricted to the subset of participants who completed both test times ($n = 44$). To examine the robustness of these analyses when the full sample ($n = 59$) was considered, we also ran individual one-way ANOVAs for reminder type independently at each level of test time. Data for all participants are displayed separately for the dominance phase (Figure 5.4), and the persistence phase (new, Figure 5.5; old, Figure 5.6).

5.1.2.3.1 Dominance

Figure 5.4 (panel A) shows that the proportion of correct responses (reporting new facts) in the dominance phase was initially quite high on the Day 3 Test across the three reminder conditions, suggesting that participants were successful at updating their semantic misconceptions with replacement facts, regardless of Day 2 memory retrieval. However, a large number of participants had faultless performance, suggesting that there could be a ceiling effect obscuring any retrieval-facilitated updating. In line with the high accuracy levels, there were relatively few intrusions across reminder conditions (reporting old facts; panel C). On the delayed test, the proportion of correct responses (panel B) had partly declined, and the proportion of intrusions (panel D) had partly increased, suggesting that misconceptions had started to recover behavioural dominance.

The reminder type by test time ANOVA with proportion of correct responses as a dependent variable confirmed that there was no significant main effect of reminder condition, $F(2, 86) = 0.33$, $p = .720$, $\eta^2_G = .00$, a significant main effect of test time, $F(1, 43) = 74.90$, $p < .001$, $\eta^2_G = .26$, and no interaction between the two factors $F(2, 86) = 0.58$, $p = .563$, $\eta^2_G = .00$.

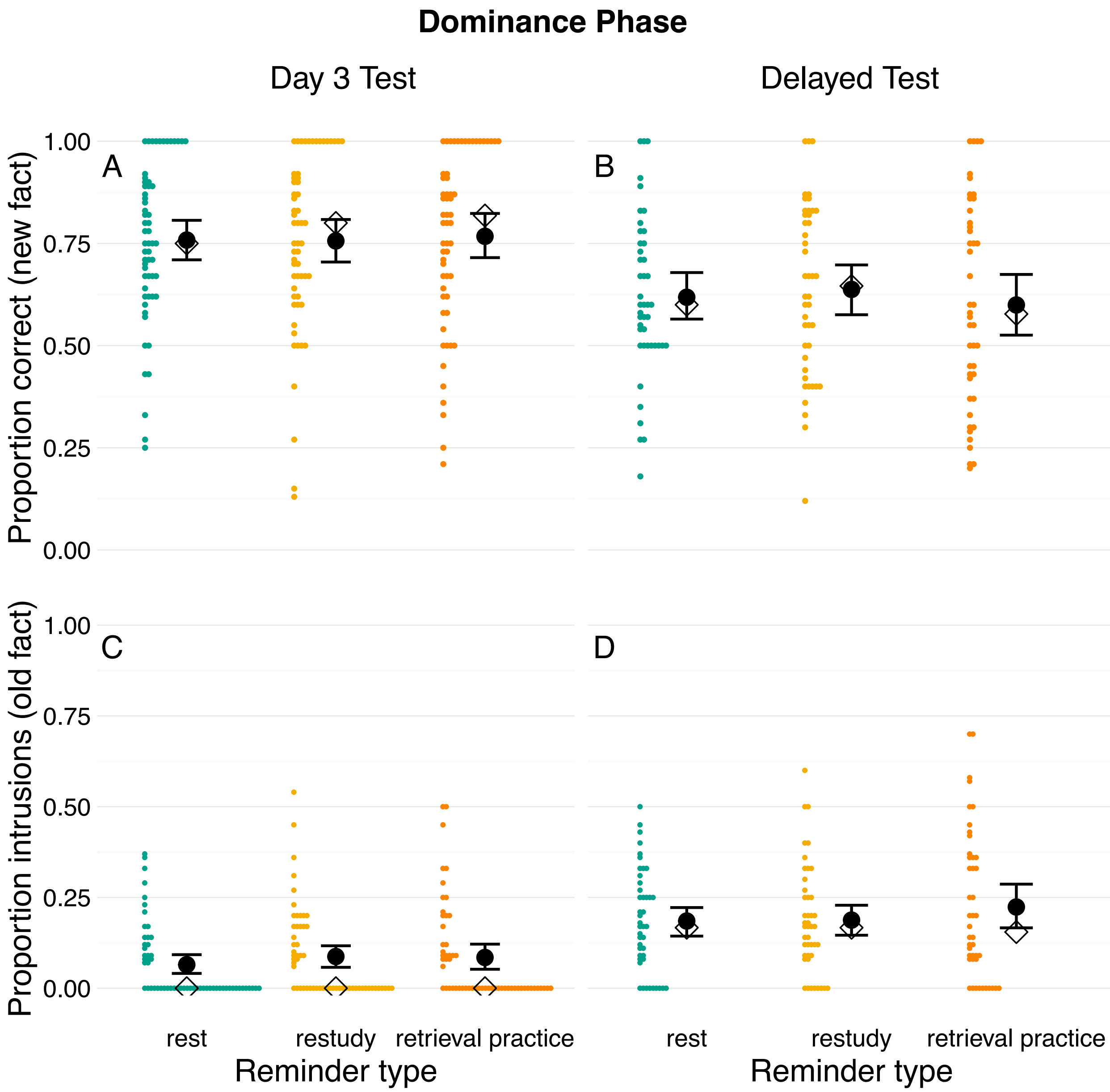


Figure 5.4: Proportion correct (Panels A, B) and proportion intrusions (Panels C, D) in the dominance phase of the Day 3 and delayed tests, as a function of reminder type (rest, restudy, retrieval practice). The dotplot shows coloured circles representing individual data points (bin width = .01). Black circles represent means and error bars indicate 95% CIs. Diamonds represent medians.

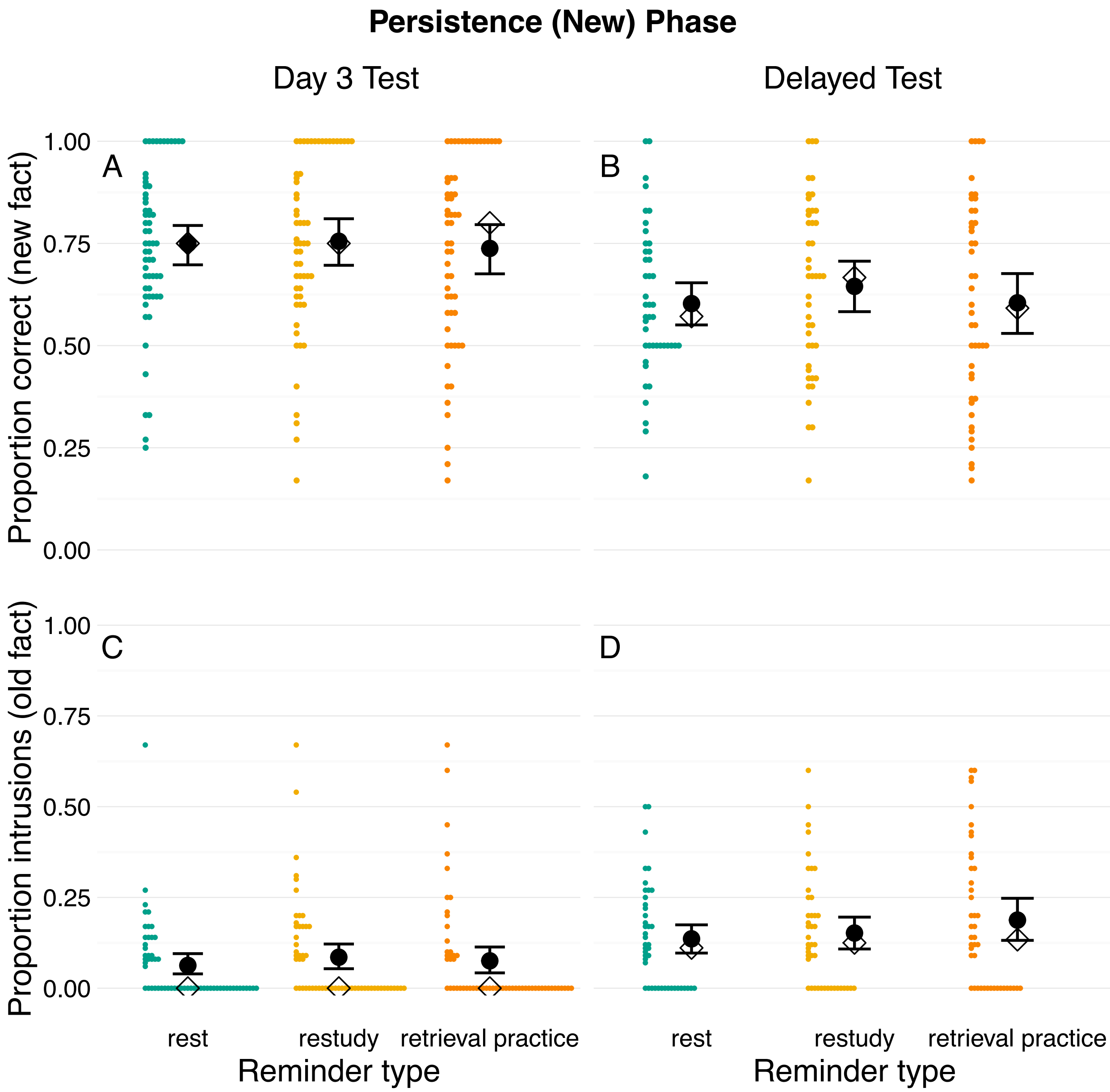


Figure 5.5: Proportion correct (Panels A, B) and proportion intrusions (Panels C, D) in the persistence (new) phase of the Day 3 and delayed tests, as a function of reminder type (rest, restudy, retrieval practice). The dotplot shows coloured circles representing individual data points (bin width = .01). Black circles represent means and error bars indicate 95% CIs. Diamonds represent medians.

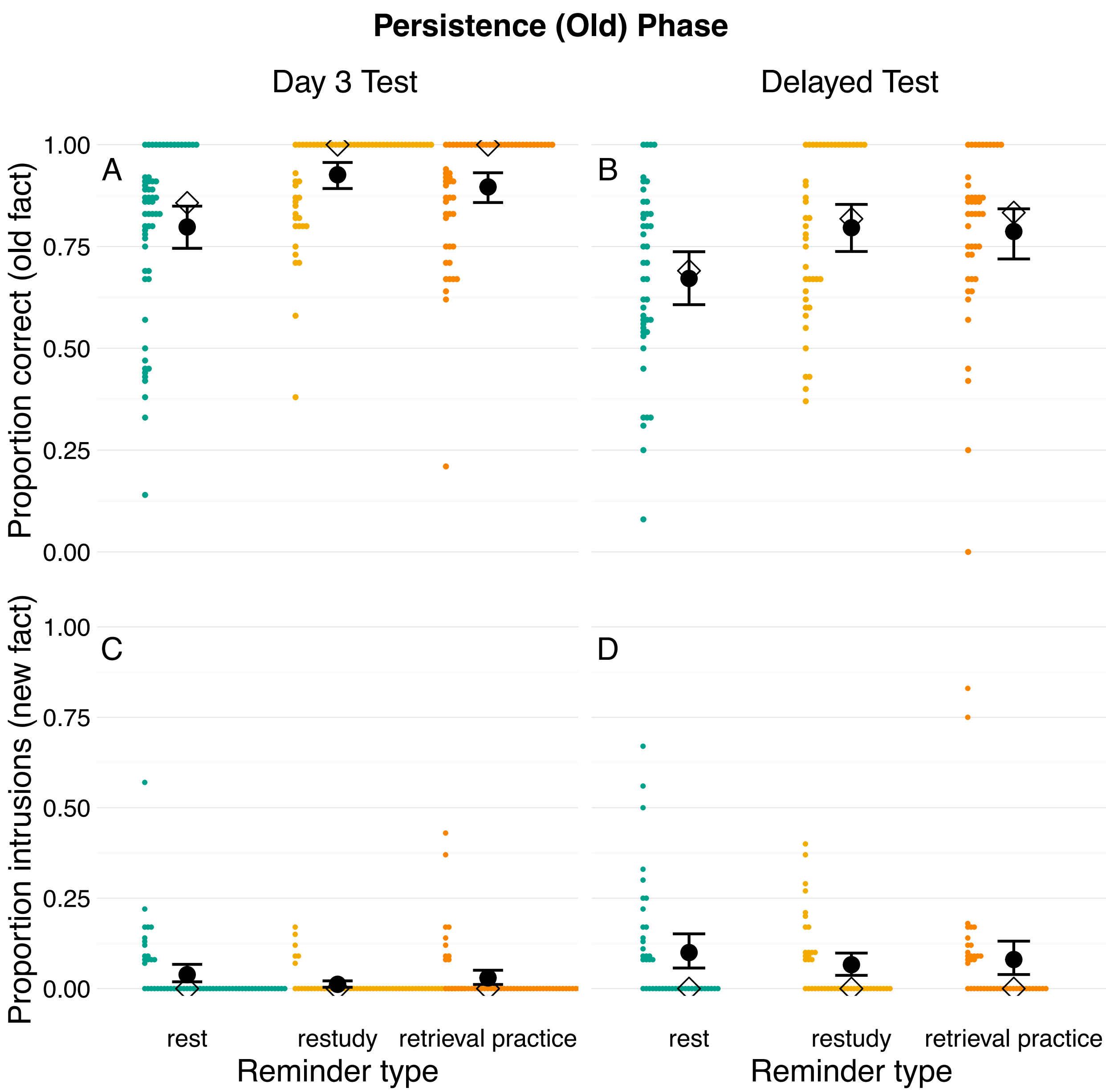


Figure 5.6: Proportion correct (Panels A, B) and proportion intrusions (Panels C, D) in the persistence (old) phase of the Day 3 and delayed tests, as a function of reminder type (rest, restudy, retrieval practice). The dotplot shows coloured circles representing individual data points (bin width = .01). Black circles represent means and error bars indicate 95% CIs. Diamonds represent medians.

ANOVAs at each level of test time employing the full sample also indicated no main effect of reminder condition on the Day 3 test, $F(2, 116) = 0.12$, $p = .889$, $\eta^2_G = .00$, or the delayed test, $F(2, 86) = 0.67$, $p = .514$, $\eta^2_G = .02$.

Similarly, with proportion of intrusions as a dependent variable, an ANOVA confirmed that there was no significant main effect of reminder condition $F(2, 86) = 1.68$, $p = .193$, $\eta^2_G = .02$, a significant main effect of test time, $F(1, 43) = 51.65$, $p < .001$, $\eta^2_G = .24$, and no interaction between the two factors, $F(2, 86) = 0.50$, $p = .608$, $\eta^2_G = .00$. ANOVAs at each level of test time employing the full sample also indicated no main effect of reminder condition on the Day 3 test, $F(2, 116) = 0.88$, $p = .418$, $\eta^2_G = .01$, or the delayed test, $F(2, 86) = 1.33$, $p = .269$, $\eta^2_G = .03$.

5.1.2.3.2 Persistence (new)

As might be expected given the high accuracy levels observed in the dominance phase, participants were very capable of reporting new facts when specifically asked to do so during the persistence phase. Figure 5.5 (panel A) shows that the proportion of correct responses (reporting new facts) was initially quite high on the Day 3 Test and had partly declined by the delayed test (panel B). There were relatively few intrusions on the Day 3 test (reporting old facts; panel C), and overall intrusion rates had partly increased by the delayed test. There did not appear to be any substantial variations across the reminder conditions for either correct responses or intrusions during either the Day 3 or delayed test.

The reminder type by test time ANOVA with proportion of correct responses as a dependent variable confirmed that there was no significant main effect of reminder condition, $F(2, 86) = 1.14$, $p = .326$, $\eta^2_G = .01$, a significant main effect of test time, $F(1, 43) = 39.15$, $p < .001$, $\eta^2_G = .19$, and no interaction between the two factors $F(2, 86) = 0.18$, $p = .837$, $\eta^2_G = .00$. ANOVAs at each level of test time employing the full sample also indicated no main effect of reminder condition on the Day 3 test, $F(2, 116) = 0.25$, $p = .777$, $\eta^2_G = .00$, or the delayed test, $F(2, 86) = 1.07$, $p = .349$, $\eta^2_G = .02$.

Similarly, with proportion of intrusions as a dependent variable, an ANOVA confirmed that there was no significant main effect of reminder condition $F(2, 86) = 2.03$, $p = .138$, $\eta^2_G = .02$, a significant main effect of test time, $F(1, 43) = 16.99$, $p < .001$, $\eta^2_G = .12$, and no interaction between the two factors, $F(2, 86) = 1.08$, $p = .345$, $\eta^2_G = .01$. ANOVAs at each level of test time employing the full sample also indicated no main effect of reminder condition on the Day 3 test, $F(2, 116) = 1.06$, $p = .351$, $\eta^2_G = .02$, or the delayed test, $F(2, 86) = 1.86$, $p = .161$, $\eta^2_G = .04$.

5.1.2.3.3 Persistence (old)

Having observed knowledge updating effects in the dominance and persistence (new) tests, it was critical to evaluate whether the old facts had been retained in the memory system. Performance in the persistence (old) phase on the Day 3 Test (Figure 5.6, Panel A) was high and appeared to be influenced by reminder condition: recall of old facts was greater in restudy and test conditions relative to the rest condition. However, as on the dominance and persistence (new) tests, there were indications of a ceiling effect. Performance declined by the delayed test (Panel B), but the pattern of higher restudy/retrieval practice performance relative to rest performance remained. Intrusion rates were very low across reminder conditions on both the Day 3 Test (Panel C) and Delayed Test (Panel D).

The reminder type by test time ANOVA with proportion of correct responses as a dependent variable indicated that there was a significant main effect of reminder condition $F(2, 86) = 17.85$, $p < .001$, $\eta^2_G = .20$ and a significant main effect of test time, $F(1, 43) = 72.42$, $p < .001$, $\eta^2_G = .21$. There was no interaction between the two factors $F(2, 86) = 0.52$, $p = .597$, $\eta^2_G = .00$. ANOVAs at each level of test time employing the full sample also indicated a significant main effect of reminder condition on the Day 3 test, $F(2, 116) = 20.02$, $p < .001$, $\eta^2_G = .26$, and the delayed test, $F(2, 86) = 9.29$, $p < .001$, $\eta^2_G = .18$.

We followed up with a series of pairwise comparisons between reminder conditions within each test stage (Table 5.2). Recall that reconsolidation theory predicts a retrieval-induced performance *decrement* (rest > retrieval practice $\geq$ restudy; H_1) whereas the hybrid testing account anticipates a retrieval-induced performance *increment* (rest < restudy < retrieval practice; H_2). Given that these two hypotheses are in opposing directions, we first employed non-directional paired t-tests to assess conventional statistical significance. As the inherent

Table 5.2: Pairwise comparisons for proportion correct in the persistence (old) phase of the Day 3 Test and Delayed Test.

Final Test	Contrast	M_{diff}[95%CI]	t	p	BF_{10} [BF_{01}]	BF_{20} [BF_{02}]
Day 3	restudy vs. rest	0.13[0.08,0.17]	5.67	<.001	0.02[66.12]	6e+04[2e-05]
Day 3	practice vs. rest	0.10[0.05,0.15]	4.07	<.001	0.03[32.75]	309.76[3e-03]
Day 3	practice vs. restudy	-0.03[-0.06,0.00]	-1.88	.065	1.42[0.71]	0.05[19.08]
Delayed	restudy vs. rest	0.12[0.06,0.19]	4.03	<.001	0.03[32.53]	274.16[4e-03]
Delayed	practice vs. rest	0.12[0.05,0.18]	3.44	.001	0.03[29.13]	49.56[0.02]
Delayed	practice vs. restudy	-0.01[-0.07,0.06]	-0.29	.775	0.18[5.53]	0.12[8.67]

Note. The unbracketed Bayes Factors (BF_{10} and BF_{20}) individually express the degree of evidence for the alternative hypotheses (H_1 and H_2) relative to the null hypothesis (H_0). The bracketed Bayes Factors (BF_{01}, BF_{02}) are simply the reciprocal of the unbracketed Bayes Factors, and individually express the degree of evidence for the null hypotheses relative to the alternative hypothesis.

limitations of orthodox statistics preclude a deeper understanding about the weight of evidence for the two competing hypotheses relative to the null hypothesis (Dienes, 2014), we employed directional Bayes Factors with 'default' JZS priors (Cauchy distribution with r scale .707; Rouder et al., 2009; Morey & Rouder, 2015) to compare each of the competing models (H_1, H_2) to a third null hypotheses (H_0) with a point null prediction (rest = restudy = retrieval practice).

These pairwise analyses clearly suggest that memory retrieval influenced performance on both the Day 3 and Delayed Tests. Performance in the restudy conditions was significantly higher than in the rest conditions and Bayesian analyses indicated decisive evidence in favour of H_2 relative to H_0. Similarly, performance in the test conditions was significantly higher than in the rest conditions, and Bayesian analysis indicated strong evidence in favour of H_2 relative to H_0. Conversely, performance in the restudy and test conditions did not differ significantly and Bayesian analysis indicated moderate evidence in favour of H_0 relative to H_2.

The intrusions ANOVA indicated that there was no significant main effect of reminder condition $F(2, 86) = 2.71$, $p = .072$, $\eta^2_G = .02$, a significant main effect of test time, $F(1, 43) =$

12.04, $p = .001$, $\eta^2_G = .12$, and no interaction between the two factors, $F(2, 86) = 0.30$, $p = .742$, $\eta^2_G = .00$. However, the individual ANOVAs at each level of test time and based on the full sample did indicate a small but significant main effect of reminder condition on the Day 3 test, $F(2, 116) = 4.00$, $p = .021$, $\eta^2_G = .06$, but not on the delayed test, $F(2, 86) = 1.68$, $p = .193$, $\eta^2_G = .04$.

5.1.3 Discussion

The goal of this study was to address whether memory retrieval would help or hinder the correction of semantic misconceptions. Both the hybrid testing account and reconsolidation theory predicted that memory retrieval would help the updating process by facilitating the learning of replacement facts. However, reconsolidation theory also predicted that erroneous 'old' information would be eliminated via a process of retrieval-induced trace overwriting (J. C. K. Chan & LaPaglia, 2013; Dudai, 2009; J. L. Lee, 2009; Nader, 2003a; Schiller & Phelps, 2011). In contrast, the hybrid testing account implied that old information would be enhanced, potentially hindering the overall goal of correcting semantic misconceptions (Carrier & Pashler, 1992; Payne, 1987; Roediger & Butler, 2011).

We began by eliciting a large number of naturally acquired semantic misconceptions (Figure 5.2 and Figure 5.3; Day 1) and then asked participants to either restudy them, retrieve them, or simply rest, prior to providing them with corrective feedback (Day 2). On Day 3, when participants were asked to respond to the questions again, there was clear evidence of knowledge updating: a shift in dominance from the old erroneous information to the new accurate information (Figure 5.4). As expected, responses on the persistence (new) test were highly similar to performance in the dominance phase (Figure 5.5). Contrary to the predictions of reconsolidation theory, and our expectations based on the hybrid testing account, the reminder manipulation had no appreciable impact on participant's performance on the Day 3 dominance test or on the persistence (new) test. Previous observations of a 'forward testing effect' (L. T. Gordon & Thomas, 2014; Pastötter & Bäuml, 2014; Szpunar, McDermott, & Roediger, 2008) suggested that retrieval practice could enhance subsequent encoding. A forward testing effect also did not emerge in the experiments conducted by

Pashler et al. (2013), as retrieval practice did not confer any additional benefit over restudy (although recall of new facts was higher overall in the retrieval practice and restudy conditions relative to the rest condition). It was possible that the use of a blocked design (retrieval practice and restudy of all items took place together before a separate block of corrective feedback) in Pashler et al. disrupted the direct connection between reminder type and new learning. By contrast, this direct connection was maintained in the present experiment by having participants perform the reminder and corrective feedback phases together for each item in turn. Therefore, it is not entirely clear why a forward-testing effect did not materialise. One account suggests that retrieval practice is effective because it helps to 'segregate' learning material, reducing any mutual interference between competing stimuli (Pastötter & Bäuml, 2014; Szpunar et al., 2008). On these grounds, the random interleaving of reminder condition (rest, restudy, retrieval practice) in the present experiment could have disrupted any attempts to effectively segregate old and new information.

In the critical persistence (old) phase, we specifically asked participants to recall the responses they had provided on Day 1 in order to assess if they had been retained in memory. The hybrid testing account suggested that old information will actually be reinforced by the reminder manipulations, particularly in the retrieval practice condition. Conversely, reconsolidation theory predicted that the old information will be overwritten, and thus eliminated. The findings on the Day 3 test clearly demonstrated that the majority of the old information had been retained (Figure 5.6), and also indicated that the reminder manipulations had been effective. Contrary to reconsolidation theory, recall of old information was *enhanced* in the reminder conditions relative to the no-reminder condition. This was more consistent with our expectations based on the hybrid testing account, however the anticipated retrieval practice vs restudy benefit did not materialise. As in the other testing phases, this could be due to a ceiling effect and will require further investigation.

The phenomenon of 'spontaneous recovery' suggests that retroactive interference effects can regress (Briggs, 1954; A. S. Brown, 1976; Wheeler, 1995), and previous studies examining semantic memory updating have reported that corrected information can begin to regain behavioural control after some delay (e.g., Butler, Fazio, & Marsh, 2011). In anticipation of such effects, we repeated the Day 3 testing regime on a delayed test approximately 3 weeks later. Consistent with these previous studies, we found that although the new replacement facts remained dominant on the delayed test, their influence had subsided and the number of ‘intrusions’ from old information had increased. As on the Day 3 Test, the reminder manipulations were only effective in the persistence (old) phase, and continued to indicate improved recall of old information in the reminder conditions relative to the no-reminder condition.

Overall, this experiment has offered some initial insight into the question of whether memory retrieval will help or hinder the correction of semantic misconceptions. Our findings suggest that effective knowledge updating can occur even in the absence of memory retrieval, contrary to claims that updating occurs within a privileged period of retrieval-induced plasticity during memory reconsolidation (J. C. K. Chan & LaPaglia, 2013; Dudai, 2009; J. L. Lee, 2009; Nader, 2003a; Schiller & Phelps, 2011). In fact, our findings indicate that retrieval could hinder the process of correction by reinforcing the erroneous information. This could offer some explanation as to why attempts to correct misconceptions can sometimes result in ‘backfire effects’, when misconceptions are reinforced rather than updated (Hart & Nisbet, 2012; Nyhan & Reifler, 2010; Seifert, 2002). Performance on the delayed test highlights that because ‘updated’ knowledge remains in the memory store, it can begin to recover its dominance after a short delay. Similarly, because the erroneous information has been retained, it could influence behaviour and decision-making in ways not identified by the memory tests used in the present investigation. The literature on the ‘continued influence effect’ suggests that ‘corrected’ information can continue to impact an individual’s judgments and decision-making, even when they readily acknowledge that it is obsolete (Bjork & Bjork, 1996; Lewandowsky et al., 2012b; Seifert, 2002).

The quotation at the outset of this chapter implied that we are adaptive creatures: while some beliefs will endure, others will be discarded when our environment signals that it is appropriate to do so. The findings reported in this chapter, add an important caveat that we will return to in Chapter 6: abandoned knowledge lingers in the memory store.

Chapter 6

The continued influence of false information stored in memory: Exploring the role of causal coverage

> "Today's news consists of aggregates of fragments. Anyone who has taken part in any event that has subsequently appeared in the news is aware of the gross disparity between the actual and the reported events."
>
> — Buckminster Fuller, 1975

Breaking news coverage is often based on sparse evidence, speculation, or hearsay. As fragmented reports are collated and verified, it may transpire that some previously disseminated information was actually incorrect. Subsequently, a retraction may be issued in order to inform individuals that they should now disregard this false information. As discussed in previous chapters, if erroneous information persists in the memory store, it has the potential to impact upon an individual's judgments, decisions, and behaviour (Bjork & Bjork, 1996; Seifert, 2002; Lewandowsky et al., 2012b). Therefore, the ideal outcome of a retraction message might be for erroneous information stored in memory to be 'edited', 'updated' or 'overwritten' (Dudai, 2009; Loftus, 1979a; Wilkes & Leatherbarrow, 1988).

Contrary to these aims, there is a sizable body of evidence indicating that individuals continue to rely on false information, even after it has been explicitly retracted (for review, see Lewandowsky et al., 2012b; Seifert, 2002). In an archetypal 'continued influence' paradigm (e.g., Ecker, Lewandowsky, Swire, & Chang, 2011b; H. M. Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988), participants are first presented with a series of short messages resembling a news report about an event. For example, in a widely used narrative, participants are told about a fire that has broken out at a warehouse. False information (e.g., "the storeroom contained paint cans and gas cylinders") is introduced early in this narrative and then explicitly retracted towards the end (e.g., "the storeroom did not contain paint cans and gas cylinders, it was actually empty"). This group of participants, who experience false information and a retraction (F+R+), is typically contrasted with a group who received the false information without a retraction (F+R-), and a group who did not experience the false information or the retraction (F-R-). On a series of subsequent inference questions (e.g., “What was a possible cause of the toxic fumes?”), the number of references to false information (e.g., “burning paint”, “the fire may have ignited the gas”) in the F+R+ group is typically lower than in the F+R- group (indicating that the retraction was effective), but does not fall to the levels of the baseline F-R- condition (indicating that the retraction was not completely effective). In other words, false information persists and continues to influence participants’ behaviour, despite them explicitly acknowledging that it has been retracted. This seems contrary to the notion that information can be routinely overwritten when it is identified as erroneous (Dudai, 2009).

The ‘real-world’ implications of the continued influence effect (from herein ‘CIE’) are potentially very serious and could have repercussions in multiple societal domains, including the media, law, and healthcare (Lewandowsky et al., 2012b). For instance, if a jury is instructed to disregard some aspect of evidence presented during a trial, it would clearly be problematic if they continued to rely on that information when considering their verdict (Fein, McCloskey, & Tomlinson, 1997). CIEs have also been observed when scientific articles are retracted (Greitemeyer, 2013), when evidence disconfirms social stereotypes (Kunda & Oleson, 1995), or when efforts are made to correct political misperceptions (Nyhan & Reifler, 2010). Attempts to derive greater insight into the phenomenon have largely been directed toward establishing the conditions under which it does and does not arise. However, the CIE has proved notoriously difficult to eradicate. The influence of false information is attenuated, but rarely eliminated, when participants are given an explicit prior warning that they will encounter false information (Ecker, Lewandowsky, & Tang, 2010), when the retraction is presented immediately after the false information rather than after a delay (H. M. Johnson & Seifert, 1994; Wilkes & Reynolds, 1999), when the retraction is repeated several times (Ecker et al., 2011b), and when the retraction is provided by a trustworthy source (Guillory & Geraci, 2013). In previous chapters, we have directly examined whether an overwriting mechanism might allow for the modification of existing knowledge. However, in this chapter, we will turn our attention to the consequences of persistent erroneous information, and examine how its influence might be mitigated by factors other than overwriting.

One theoretical interpretation of the CIE is based on the assumption that participants form internal cognitive representations called ‘mental models’ which characterise their understanding of the event (H. M. Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988; Wilkes & Reynolds, 1999). The model is thought to represent a ‘causal chain’, or network of associative connections, that link the false information with other details about the event. For example, in the warehouse fire story, participants appear to form causal links between the volatile materials and other *auxiliary details* reported in the narrative, such as the presence of “oily smoke and sheets of flame”, “intense heat”, “explosions”, and “toxic fumes”. This is clearly evident in participants’ responses to the inference questions (e.g., “Why do you think the fire was particularly intense?” Answer: “The pressurized cylinders”; Seifert, 2002). The intention of a retraction message is that participants should ‘edit out’ the false information (Wilkes & Leatherbarrow, 1988). However, individuals appear to prefer mental models in which information is mutually consistent, or ‘coherent’ (Johnson-Laird, 2012). Removing the false information would leave a ‘causal gap’, and disrupt the coherence of their model. This could explain why participants appear to continue to rely on a model based on false information, despite knowing that a central link in the causal chain has been broken (H. M. Johnson & Seifert, 1994; Seifert, 2002; Wilkes & Reynolds, 1999).

The mental models theory might also offer some insight into why one of the most effective techniques for reducing the CIE is the provision of an alternative explanation for the event (Ecker et al., 2010; Ecker, Lewandowsky, & Apai, 2011a; H. M. Johnson & Seifert, 1994; Wilkes & Reynolds, 1999). For example, in H. M. Johnson & Seifert (1994; warehouse fire narrative), when the retraction statement about the paint cans and gas cylinders was accompanied by an alternative explanation suggesting that arson-related materials had been discovered, this was more effective at reducing CIE than providing a retraction on its own. An alternative explanation might ‘fill the causal gap’, thereby maintaining the model’s coherence, and reducing the need to rely on the false information (H. M. Johnson & Seifert, 1994; Seifert, 2002). Critically, this could also explain why alternative explanations do not necessarily eliminate the CIE, and why different alternative explanations vary in their efficacy. Some alternative explanations will provide better replacements for the false information because they are able to explain more features of the original news report, a property referred to as 'causal coverage' (Seifert, 2002).

Consider that in the traditional warehouse fire narrative, the auxiliary details tend to reinforce an explanatory theme that is consistent with the false information (from herein, the 'false theme'). For example, "oily smoke and sheets of flame", "explosions", and "toxic fumes", are all consistent with the presence of "paint cans and gas cylinders". When asked questions such as, "What could have caused the explosions?" participants in the retraction group (F+R+) might continue to rely on a mental model based on false information because it provides the most extensive causal coverage of the event compared to any alternative account they can think of. Although participants appear to be able to generate their own explanations for the event when asked to do so (e.g., "cigarettes", "gas leak", "lightening"), this does not appear to be an effective method for reducing the CIE (H. M. Johnson & Seifert, 1999). It could be that these self-generated alternative accounts have relatively low causal coverage, only explaining one or two features of the narrative whereas the false information provides a plausible account of multiple features.

Similarly, when researchers provide participants with an alternative account (from herein, 'alternative theme'), it could be more or less effective at replacing the false information (and therefore eliminating CIE), because of its relative causal coverage. In H. M. Johnson & Seifert (1994) for example, the 'arson' alternative explanation might provide a reasonable account for the fire, but perhaps does not explain the "oily smoke and sheets of flame", "explosions", and "toxic fumes" as effectively as the presence of paint cans and gas cylinders, which are inherently volatile. Similarly, in Ecker et al. (2010), participants read a narrative about a minibus crash and were falsely informed that the passengers were a group of elderly nursing home residents. Additional auxiliary details about the event appear consistent with this false theme: "The rescue crew also reported difficulty in getting both injured and uninjured passengers out of the minibus even though the exits were clear", "Local television shows live footage of uninjured passengers having problems getting up the embankment", and "Rescue crew can be heard remarking that the uninjured passengers were unable to help in the rescue efforts." By contrast, the alternative account, offered alongside the retraction statement, seems to have relatively low causal coverage, "A second statement from the police has stated that the passengers on the minibus were not elderly people but college hockey players returning from a victory party after the state hockey final." Therefore, although provision of an alternative can attenuate the CIE, it may not always be completely effective because the false theme has greater causal coverage than the alternative theme.

In a naturally occurring 'breaking news' situation, one might expect that auxiliary details reported about an event will tend to cohere around what actually happened, i.e., an 'alternative theme', rather than a hypothetical scenario based on a single piece of false information (assuming that the news agency does not have some malign intent to mislead). Consider for example, a breaking news report about a plane crash. Soon after the crash, the news agency receives an anonymous tip-off and announces that the pilot was drunk before she boarded the flight. The organisation continues reporting 'auxiliary details' about the event as information becomes available: firstly, an eye witness reports seeing flames coming from one of the aircraft's jet engines; secondly, a reporter gets his hands on the flight checklist, which appears to indicate that there were some problems with the engine before take-off; and thirdly, it is discovered that the chief mechanic is currently on holiday, leaving his inexperienced apprentice in charge of repairs. Clearly, the false information in this scenario has fairly low causal coverage and a plausible, coherent, alternative is readily available: the plane crashed because of a mechanical problem with one of its engines. A retraction message stating that, for example, the anonymous tip-off actually came from a disgruntled ex-employee of the airline, would leave very little reason to rely on the false information. It seems intuitive that the availability of a plausible and coherent alternative should reduce the influence of the false information, especially in conjunction with a retraction of the false information.

Although a number of different alternative accounts have been employed in CIE studies with varying degrees of success (Ecker et al., 2010, 2011a; H. M. Johnson & Seifert, 1994; Wilkes & Reynolds, 1999), we are not aware of any systematic attempt to evaluate the causal coverage hypothesis. Here we report an experiment that examined whether the CIE is eliminated when an alternative theme affords greater causal coverage of the event compared to the false

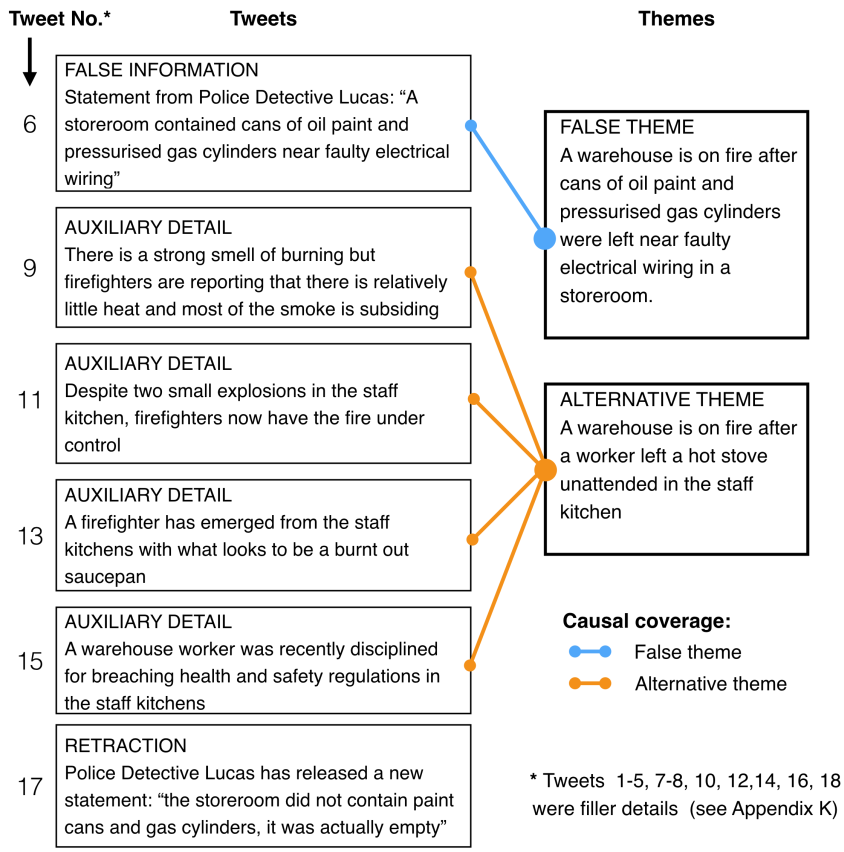


Figure 6.1: Key structural elements of the warehouse fire narrative. The alternative theme has greater causal coverage than the false theme. Note that Tweets 6 and 17 varied between experiment groups (F-R-, F+R-, F+R+) and filler tweets are not shown (see Appendix K for details).

theme. To achieve this, we adjusted the core structure of the event narrative, such that the auxiliary details (e.g., "A warehouse worker was recently disciplined for breaching health and safety regulations in the staff kitchens") supported an alternative theme more than the false theme (Figure 6.1). We did not explicitly inform participants about the alternative theme, but allowed them to arrive at their own conclusions, as might occur in a real world breaking news situation. We anticipated that individuals would seek an explanatory account of the event that maximised its causal coverage of the auxiliary details (Seifert, 2002), and therefore expected that following retraction of the false information, the CIE would be eliminated.

We also took the opportunity to update the presentation format of the experimental materials, delivering the narrative as a series of 'tweets' (see Figure 6.2) as used on the social media

Figure 6.2: Example tweet used in Experiment 12.

platform Twitter. The spread of misinformation on social media is increasingly problematic (Del Vicario et al., 2016). For example, during an unprecedented bout of civil unrest in London (UK) in August 2011, rumours began to circulate on Twitter suggesting that a tiger was on the loose in the district of Primrose Hill after rioters had liberated it from London Zoo, the army had been deployed in the district of Bank, and the London Eye was on fire (Procter, Vis, & Voss, 2011). Social media presents unique challenges because the mechanisms of self-correction are less well defined and information can proliferate extremely rapidly. Although our initial experiment closely follows the design of previous CIE studies, there is scope for extending the Twitter-based paradigm we have developed here to address the unique challenges of misinformation correction on social media.

To enhance the generalisability of our findings, we developed two different narratives that had the same core structure (see Figure 6.1 and Appendix K for details) but different superficial features. One narrative was the aforementioned warehouse fire study, and the second narrative described the aftermath of a 'zoo breakout' in which a group of monkeys escaped after a work experience student left their cage door open (false theme) or after an animal rights activist cut a hole in the fence of their enclosure (alternative theme). Finally, we employed an additional test phase in order to evaluate if the retraction had caused memory impairment. In the archetypal CIE paradigm, higher performance in F+R+ vs F-R- indicates the continued influence of false information. But the drop in performance from F+R- to F+R+ suggests that the retraction has at least been partly successful at updating participants' beliefs. Consequently, we employed a forced-choice 'modified' recognition test (see Chapter 4; M. McCloskey & Zaragoza, 1985a) to assess whether false information remains stored in memory even if it no longer influences inferential reasoning about the event. If the information has persisted, this would preclude the role of an overwriting or ‘destructive updating’ mechanism (Loftus, 1979a).

In summary, the specific hypotheses under scrutiny were as follows:

Manipulation check (influence of false information)

The number of references to false information will be significantly higher in a false information without retraction group (F+R-) relative to a no false information and no retraction baseline group (F-R-), indicating that presence of false information successfully induced participants to rely on the false theme (i.e., F+R- > F-R-).

Support for the null hypothesis (F+R- = F-R-) would indicate that the false information was not effective so it would not be possible to assess the impact of a retraction.

Hypothesis 1 (influence of retraction)

The number of references to false information will be significantly lower in a false information with retraction group (F+R+) compared to a false information without retraction group (F+R-), indicating that the retraction successfully reduced the influence of the false information (i.e., F+R+ < F+R-).

Support for the null hypothesis (F+R+ = F+R-) would indicate that the retraction had no impact at all, which would be a surprisingly strong CIE.

Hypothesis 2 (continued influence effect)

The number of references to false information will be greater in a false information with retraction group (FI+R+) relative to a no false information and no retraction baseline group (FI-R-), indicating a CIE effect (i.e., F+R+ > F-R-).

Support for the null hypothesis (F+R+ = F-R-) would indicate that the CIE has been eliminated.

Hypothesis 3 (recognition accuracy)

(a) Accuracy on the forced-choice recognition test will be significantly higher in the retracted false information group (F+R+) relative to the no false information and no retraction baseline group (F-R-), indicating the persistence of at least some false information in memory despite the retraction (i.e., F+R+ > F-R-).

Support for the null hypothesis (F+R+ = F-R-) would be indicative of memory impairment.

(b) Accuracy on the forced-choice recognition test will be comparable between the retracted false information group (F+R+) and the false information without retraction group (F+R-), indicating the absence of any storage-based memory impairment caused by the retraction (i.e., F+R+ = F+R-).

Support for the alternative hypothesis (F+R+ < F+R-) would be indicative of memory impairment.

6.1 Experiment 12

6.1.1 Methods

The hypotheses, methods, and analysis plan for this study were pre-registered (http://dx.doi.org/10.17605/osf.io/x2rzh)[44]. All experimental materials, data, and analysis code pertaining to this chapter have been made publicly available on the Open Science Framework (https://osf.io/jc4mw). All data exclusions, manipulations, and measures conducted during this study are reported.

6.1.1.1 Participants

6.1.1.1.1 Power-analysis

A power analysis using an effect size for the critical F+R+ vs. F-R- contrast observed in previous research (contrast 0 in Table 1 of Ecker et al., 2011b) suggested that 66 participants would be required in each condition to achieve reasonable statistical power (d = 0.49, α = 0.05, 1-β = 0.80). Therefore, we intended to recruit 198 participants.

[44]Note that the hypotheses and analyses used here apply only to the first experiment of a planned series and have therefore been slightly modified from the pre-registration.

6.1.1.1.2 Stopping rule

We pre-registered a 'stopping rule' to constrain researcher degrees of freedom in the data collection process (Simmons et al., 2011). Data collection was to be terminated upon reaching the sample size specified in the power analysis, unless that target had not been reached by 28/02/2016. We slightly exceeded ($n = 204$) our sample size target ($n = 198$) as 6 additional participants were needed in order to ensure equal counter-balancing conditions.

6.1.1.1.3 Sample demographics

Two-hundred and four participants (114 female; 90 male; age: $Mdn = 25$ years, range = 18-60 years) were recruited from the University College London (UCL) multi-occupation subject pool ($n = 72$; https://uclpsychology.sona-systems.com/) and the online crowd-sourcing platform Prolific Academic ($n = 132$; http://www.prolific.ac/). The sample recruited from UCL attended our testing rooms in person and the online sample completed the study remotely on their own personal computing devices. All participants reported that they spoke fluent English and were between 18-65 years of age. They also completed all stages of the study, finished the study within the 60 minute maximum time limit, and passed all instructional and attentional checks (see Appendix I). These inclusion criteria were pre-registered, and all data from participants who failed to meet these criteria were excluded and replaced. Participants provided informed consent and received monetary compensation or course credits. The local UCL ethics committee approved the study.

6.1.1.2 Design

The experiment employed a between-subjects design with three groups: false information with retraction (F+R+), false information without retraction (F+R-), and no false information or retraction (F-R-). We also employed two narratives with the same underlying structure but different superficial features (warehouse fire and zoo breakout, see Materials for details) which were counter-balanced. Thus the design had six cells: two narrative conditions, each with three groups. Participants were randomly allocated to these cells with the contingency that all final cell sizes were balanced ($n = 34$ per cell).

The two primary dependent measures were references to the false theme and references to the alternative theme on the inference questions. We also measured participants' recall of filler items, their awareness of a retraction message, and their recognition accuracy (i.e., 'hit rate') for the false information.

Participants were not made aware of the explicit purpose of the study or which experimental group they had been assigned to until the debriefing stage. Researchers involved in data collection were aware of the purpose of the study, but were not aware of which groups participants had been assigned to until after data collection was complete (group assignment was handled by a computer algorithm). Group labels were also hidden during response coding (see below).

6.1.1.3 Materials

Two narratives were developed, each with a similar underlying structure (see Appendix K) but rather different superficial features. The 'warehouse fire' was adapted from a similar narrative often used in research on the CIE (e.g., Ecker et al., 2011b; H. M. Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988), and the 'zoo breakout' was novel. Each narrative consisted of 18 short messages (each less than 140 characters) resembling individual 'Tweets' from the social media platform Twitter (e.g., Figure 6.2). All messages appeared to come from a (fictional) news outlet called 'News Today'. Both narratives elicited the CIE in a pilot study ($n = 39$).

Slightly modified versions of the warehouse and zoo narratives were employed depending on experimental group (F+R+, F+R-, F-R-). The 6th message either contained false information (e.g., *Statement from Police Detective Lucas: "A storeroom contained cans of oil paint and pressurised gas cylinders near faulty electrical wiring"*) or did not (*Statement from Police Detective Lucas: "We are working with the fire services to rapidly get the situation under control"*). The 17th message either contained a retraction message (e.g., *Police Detective Lucas has released a new statement: "the storeroom did not contain paint cans and gas* *cylinders, it was actually empty")* or did not (e.g., *Police Detective Lucas has released a new statement: "the fire is now under control and the area will soon be declared safe")*.

The other messages were the same across experimental groups and either referred to general non-causal information about the event (12x 'filler' details; e.g., *A small crowd of onlookers has gathered on a nearby pedestrian bridge to watch as the fire consumes the warehouse*) or referred to information that was specifically related to the cause of the event (4x 'auxiliary details'; e.g., *A warehouse worker was recently disciplined for breaching health and safety regulations in the staff kitchens*).

The critical departure from previous continued influence studies was to ensure that auxiliary details in the narratives preferentially supported the alternative theme (Figure 6.1; see Appendix K for details). In the warehouse narrative, all of the auxiliary details implied that the fire was caused by a worker leaving a hot stove unattended in the staff kitchen, even though the false information suggested that the fire was caused by cans of oil paint and pressurized gas cylinders left near faulty electrical wiring in a storeroom. In the zoo narrative, all of the auxiliary details implied that the monkeys had escaped after an animal rights activist cut a hole in the fence of their enclosure even though the false information suggested that a work experience student had left their cage door open.

6.1.1.4 Procedure

Participants were told that they were taking part in a study investigating how well people remember news information presented on the social media platform Twitter. They read a short paragraph of instructions (Appendix J) and then responded to an instruction-attention check (Appendix I),

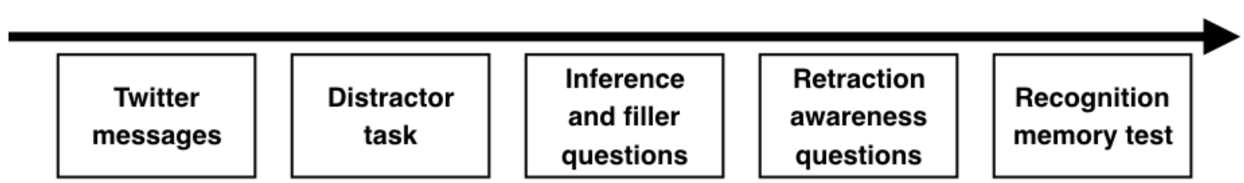


Figure 6.3: Stages of the continued influence paradigm used in Experiment 12.

The core procedure is shown in Figure 6.3. Participants were presented with a narrative (Appendix K) in the form of a series of sequential Twitter messages (‘Tweets’; e.g., see Figure 6.2). Depending on the participant’s experimental group, the narrative either contained false information (F+R-, F+R+) or did not (F-R-), and either contained a retraction message (F+R+) or did not (F+R-, F-R-). Messages appeared one at a time for a minimum of 5s each, and no maximum time. A button had to be clicked in order to proceed to the next message. Participants were unable to go back and view previous messages. After all of the tweets had been presented, participants completed a 5 minute distractor task consisting of arithmetic questions before moving on to the testing stages.

During the test stages, participants encountered separate blocks of questions presented in a pre-defined order: inference questions (x8) and filler questions (x8; intermixed), retraction awareness questions (x2), and a recognition test (x1). The order of questions within each block was randomized.

In the first block of testing they were asked questions about the event described in the tweets. Some of these questions (‘inference questions’, Appendix M: Table M.1 and Table M.2) required them to reason about causes of the event and go beyond the information provided in the tweets. For example, “What was the possible cause of the strong burning smell?” Other questions (‘filler questions’, Appendix M: Table M.3 and Table M.4) were more general and did not have a direct bearing on the causes of the event. For example, “Where did onlookers gather to watch the fire?” A second attention check was embedded within this set of questions (see Appendix I).

In the second block of testing, participants answered questions that assessed their awareness of the retraction message (see Appendix M). For example, “Were you aware of any corrections or contradictions in the messages that you read?”

In the final block of testing, participants were asked to complete a two-alternative forced choice recognition memory test (also known as the “Modified Test”; see Chapter 4; M. McCloskey & Zaragoza, 1985a) to assess the persistence of false information in the memory store. This involved presentation of the false information tweet in its original form vertically adjacent to a second message that contained novel information instead of the false information (a lure;

Appendix L). The spatial position (top or bottom) of the tweets was counter-balanced.

6.1.1.5 Response coding

Responses were coded according to the pre-registered coding framework (see Appendix M). In summary, for the inference questions, coders classified responses as references to the false theme, the alternative theme, or neither. References had to be ‘causal’ in nature (see Ecker et al., 2011a): the participant had to explicitly state, or strongly imply, that the warehouse fire was caused by gas and oil paints/unattended hot stove or that the zoo breakout was caused by the cage being left open by a work experience student/an activist cutting a hole in the enclosure.

- References related to neither theme received a score of zero on both dependent measures.
- References to the false theme scored 1 on the ‘references to false information’ measure.
- References to the alternative theme scored 1 on the ‘references to alternative theme’ measure.
- If the participant made references to both themes they scored 0.5 on each measure.

The maximum score that could be achieved for each measure on each question was therefore 1, even if multiple references were made in the same response. Consequently, the maximum score for each dependent measure for all inference questions was 8.

10% of the responses ($n = 20$) were joint-coded to assess the reliability of the coding framework. As high inter-rater agreement was obtained (96%), the remaining responses were single coded.

6.1.2 Results

Statistical significance was defined at the .05 level. All Bayes Factors (BF_{10}; Rouder et al., 2009) indicate the degree of evidence for the alternative hypothesis (H_1) relative to the null hypothesis (H_0) and use a ‘default’ JZS prior (Cauchy distribution with scale r = .707).

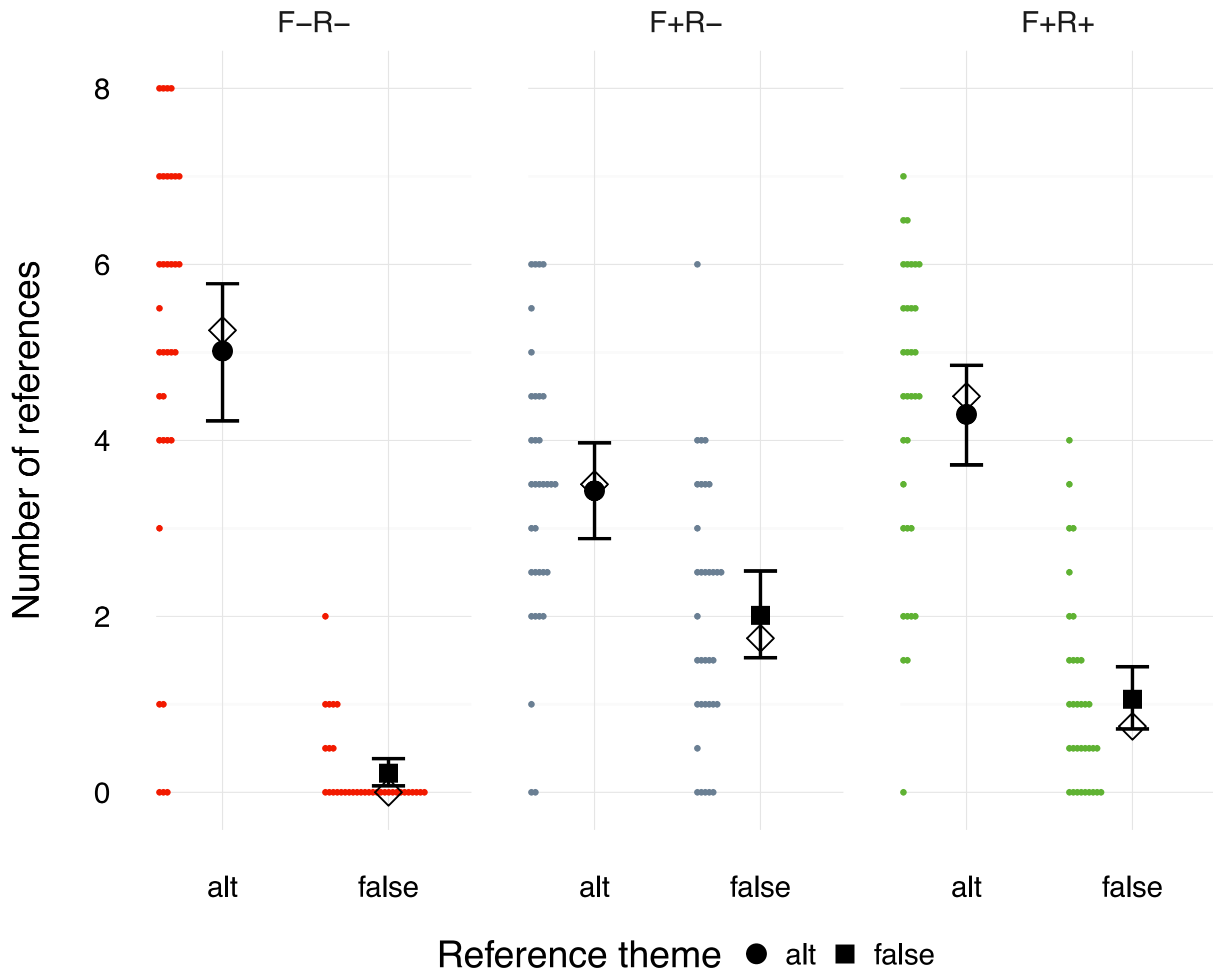


Figure 6.4: Zoo narrative condition. Number of references to the false theme and the alternative (alt) theme based on experimental group: No false information or retraction (F-R-), false information without retraction (F+R-), and false information with retraction (F+R+). Maximum number of references to the each theme was 8. Dotplots show coloured circles representing indvidual data points (bin width = .05). Means are represented by black circles (false theme) or black squares (alternative theme), and error bars represent 95% CIs. Diamonds represent medians.

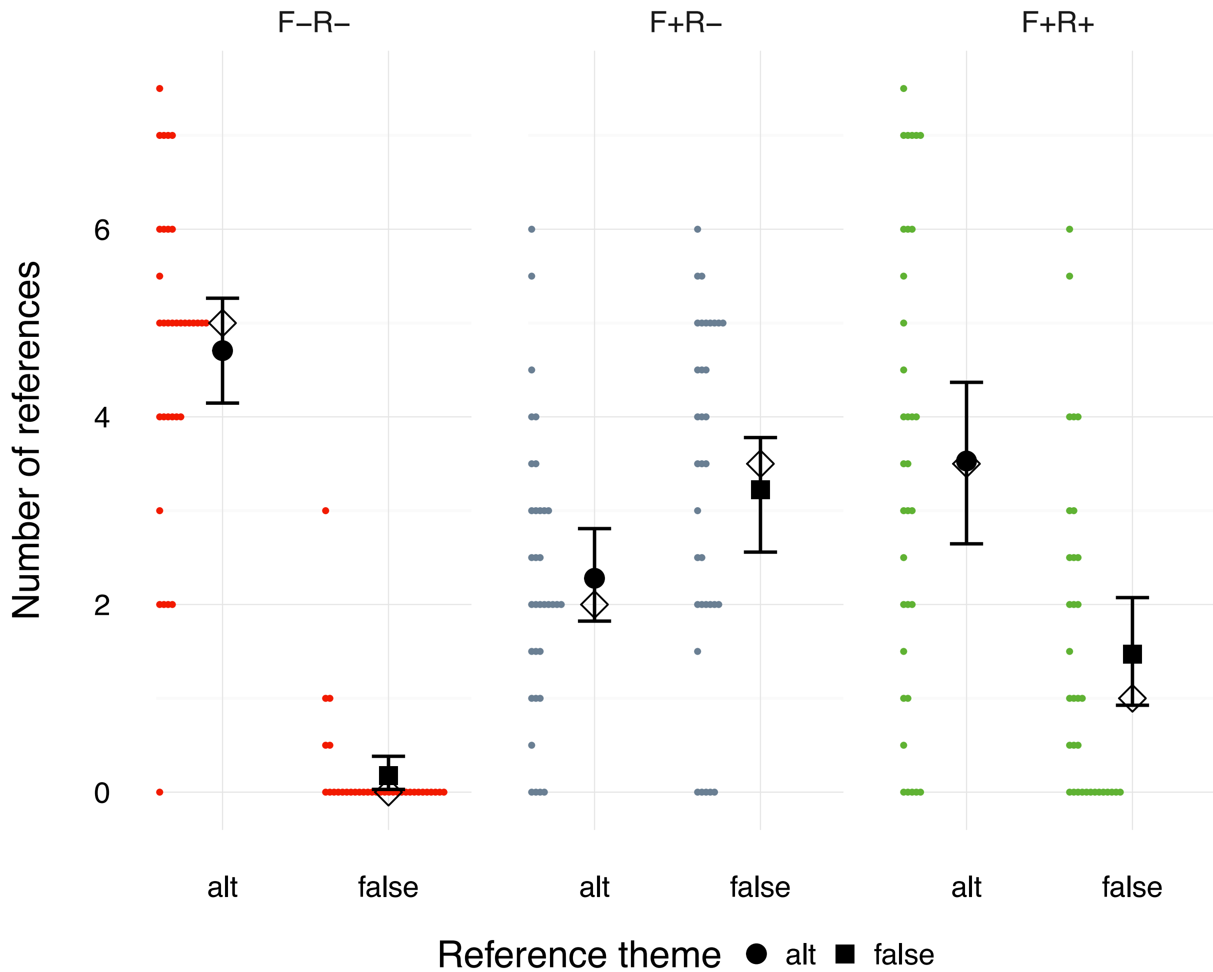


Figure 6.5: Warehouse narrative condition. Number of references to the false theme and the alternative (alt) theme based on experimental group: No false information or retraction (F-R-), false information without retraction (F+R-), and false information with retraction (F+R+). Maximum number of references to the each theme was 8. Dotplots show coloured circles representing indvidual data points (bin width = .05). Means are represented by black circles (false theme) or black squares (alternative theme), and error bars represent 95% CIs. Diamonds represent medians.

6.1.2.1 Inference questions

A 2 x 3 between-subjects ANOVA with narrative type (Zoo, Warehouse) and experimental group (F+R-, F+R+, F-R-) as independent variables and references to the false theme as a dependent variable, indicated a significant main effect of experimental group $F(2, 198) = 59.12$, $p < .001$, $\eta^2_G = .37$, a significant main effect of narrative type, $F(1, 198) = 8.30$, $p = .004$, $\eta^2_G = .04$ and an interaction between the two factors, $F(2, 198) = 4.02$, $p = .019$, $\eta^2_G = .04$. We therefore proceeded with separate analyses for the zoo narrative (Figure 6.4) and warehouse narrative (Figure 6.5).

A one-way between-subjects ANOVA with references to the false theme as a dependent variable, indicated a significant main effect of experimental group for both the zoo narrative, $F(2, 99) = 23.44$, $p < .001$, $\eta^2_G = .32$, and the warehouse narrative, $F(2, 99) = 35.87$, $p < .001$, $\eta^2_G = .42$. A similar ANOVA for references to the alternative theme indicated a significant main effect of experimental group for the zoo narrative, $F(2, 99) = 5.86$, $p = .004$, $\eta^2_G = .11$, and the warehouse narrative, $F(2, 99) = 13.54$, $p < .001$, $\eta^2_G = .21$. We therefore proceeded with pairwise contrasts between all experimental groups within the zoo (Table 6.1, contrasts 1-6) and warehouse (Table 6.2, contrasts 7-12) conditions for both primary dependent variables (references to false theme and alternative theme). Within-group contrasts between reference themes are shown in Table 6.3 (contrasts 13-18). To aid identification in the following text, all contrasts are referred to via the number assigned to them in the leftmost column of each table. The Holm-Bonferroni procedure was used to correct p-values for multiple-comparisons (see table captions for the different test families). As the overall pattern of findings was similar for the zoo narrative (Figure 6.4) and warehouse narrative (Figure 6.5), they are discussed together below, with any salient differences explicitly noted.

6.1.2.1.1 Baseline performance (F-R-)

As expected, the baseline F-R- conditions (from which false information was absent) yielded an extremely low number of spontaneous false theme references. Conversely, there was a relatively high number of references to the alternative theme, suggesting that the auxiliary

Table 6.1: Between-group pairwise contrasts for zoo narrative by reference theme (False, Alternative).

No.	Reference Theme	Contrast/Prediction	M_{diff}[95%CI]	t	p	BF_{10}
1	False	F+R- > F-R-	1.79[1.27,2.32]	6.88	<.001	6e+06
2	False	F+R+ > F-R-	0.84[0.43,1.25]	4.12	<.001	414.48
3	False	F+R+ < F+R-	-0.96[-1.58,-0.33]	-3.07	.002	23.91
4	Alternative	F+R- <> F-R-	-1.59[-2.56,-0.62]	-3.28	.005	20.15
5	Alternative	F+R+ <> F-R-	-0.72[-1.72,0.28]	-1.44	.156	0.60
6	Alternative	F+R+ <> F+R-	0.87[0.06,1.67]	2.16	.069	1.75

Note. *F-R-*, No false information or retraction; *F+R-*, false information without retraction; *F+R+* false information with retraction; BF_{10}, Bayes Factor indicating the degree of evidence for the alternative hypothesis relative to H_0; p, Holm-Bonferroni corrected p-value (test families by reference theme).

Table 6.2: Between-group pairwise contrasts for warehouse narrative by reference theme (False, Alternative).

No.	Reference Theme	Contrast/Prediction	M_{diff}[95%CI]	t	p	BF_{10}
7	False	F+R- > F-R-	3.04[2.38,3.71]	9.21	<.001	6e+10
8	False	F+R+ > F-R-	1.29[0.67,1.92]	4.19	<.001	518.26
9	False	F+R+ < F+R-	-1.75[-2.61,-0.89]	-4.06	<.001	346.80
10	Alternative	F+R- <> F-R-	-2.43[-3.20,-1.66]	-6.30	<.001	4e+05
11	Alternative	F+R+ <> F-R-	-1.18[-2.20,-0.15]	-2.30	.028	2.27
12	Alternative	F+R+ <> F+R-	1.25[0.26,2.24]	2.54	.028	3.65

Note. *F-R-*, No false information or retraction; *F+R-*, false information without retraction; *F+R+* false information with retraction; BF_{10}, Bayes Factor indicating the degree of evidence for the alternative hypothesis relative to H_0; p, Holm-Bonferroni corrected p-value (test families by reference theme).

Table 6.3: Within-group pairwise contrasts between false theme and alternative theme references.

No.	Narrative	Group	Contrast/Prediction	M_{diff}[95%CI]	t	p	BF_{10}
13	Warehouse	F-R-	Alternative <> False	4.53[3.84,5.22]	13.39	<.001	1e+12
14	Zoo	F-R-	Alternative <> False	4.79[3.91,5.67]	11.09	<.001	7e+09
15	Warehouse	F+R-	Alternative <> False	-0.94[-1.98,0.10]	-1.85	.074	0.84
16	Zoo	F+R-	Alternative <> False	1.41[0.54,2.28]	3.29	.007	14.96
17	Warehouse	F+R+	Alternative <> False	2.06[0.73,3.39]	3.14	.007	10.55
18	Zoo	F+R+	Alternative <> False	3.24[2.38,4.09]	7.70	<.001	2e+06

Note. *F-R-*, No false information or retraction; *F+R-*, false information without retraction; *F+R+* false information with retraction; BF_{10}, Bayes Factor indicating the degree of evidence for the alternative hypothesis relative to H_0; p, Holm-Bonferroni corrected p-value (test families by narrative).

narrative details provided a plausible alternative explanation of events. These differences were statistically significant and yielded strong evidence in favour of the alternative hypothesis according to a Bayes Factor (contrasts 13 and 14).

6.1.2.1.2 Influence of false information (Manipulation check, F+R- vs. F-R-)

When false information was introduced, but not retracted (F+R-), the number of false references increased (contrasts 1 and 7) and the number of alternative references decreased (contrasts 4 and 10). All contrasts were statistically significant and yielded strong evidence in favour of the alternative hypothesis according to a Bayes Factor.

The within-group comparisons of alternative vs false references showed a different pattern for the zoo and warehouse narratives. In the zoo condition, there were more alternative references relative to false references, a statistically significant difference accompanied by a strong evidence in favour of the alternative hypothesis according to a Bayes Factor (contrast 16). Conversely, in the warehouse condition there were more false references than alternative references, but the difference was not statistically significant and yielded an inconclusive Bayes Factor (contrast 15). Therefore, the false information did not appear sufficient to displace the alternative theme as the dominant explanatory account for the event in the zoo narrative. The situation was more ambiguous for the warehouse narrative.

6.1.2.1.3 Influence of retraction (Hypothesis 1, F+R+ vs. F+R-)

In the group in which false information was retracted (F+R+), there was a decrease in the number of false references (contrasts 3 and 9) and an increase in the number of alternative references (contrasts 6 and 12). There were substantially more alternative references than false references (contrasts 17 and 18). All contrasts were statistically significant. The reduction in false references yielded strong evidence in favour of the alternative hypothesis according to a Bayes Factor. The increase in alternative references was less robust but still yielded moderate evidence in favour of the alternative hypothesis. Therefore, the retraction message reduced reliance on the false information as predicted and correspondingly led to increased reliance on the alternative theme.

6.1.2.1.4 Continued influence effect (Hypothesis 2, F+R+ vs. F-R-)

The critical question with regards to the CIE was whether the retraction was not just capable of reducing the effect of the false information (see above), but eliminating its impact entirely. This required a comparison between the baseline condition (F-R-) and the retraction condition (F+R+). A CIE was obtained: despite the retraction, the number of false references was higher than in the baseline condition (contrasts 2 and 8). These differences were statistically significant and yielded very strong evidence in favour of the alternative hypothesis according to Bayes Factors.

The number of alternative theme references was slightly lower than in the baseline condition (contrasts 5 and 11) although these differences were not statistically significant and Bayes Factors indicated that the data were fairly inconclusive. Overall, it is not clear if the retraction was effective at fully restoring reliance on the alternative theme.

Table 6.4: Descriptive statistics for filler scores, retraction awareness scores, and recognition scores, across experimental groups.

Group	Filler $M(SD)$	Retraction Awareness $M(SD)$	Recognition $M(SD)$
F-R-	4.96(1.82)	0.07(0.31)	0.51(0.50)
F+R-	5.15(1.93)	0.40(0.52)	0.96(0.21)
F+R+	5.28(1.84)	1.44(0.72)	0.96(0.21)

6.1.2.2 Filler scores

Responses to the filler questions were broadly similar across experimental groups (Table 6.4). A 2 x 3 between-subjects ANOVA with experimental group (F+R-, F+R+, F-R-) as an independent variable and filler scores as a dependent variable, indicated no significant main effect of experimental group, $F(2, 201) = 0.52$, $p = .596$, $\eta^2_G = .01$.

6.1.2.3 Retraction Awareness

Retraction awareness scores suggested that participants noticed and understood the retraction message when it was present (in F+R+; Table 6.4). Excluding participants from the F+R+ group who failed to noticed the retraction (i.e., awareness score = 0; $n = 9$) did not have a substantial impact on the overall pattern of results.

6.1.2.4 Recognition (Hypothesis 3)

In a final recognition test participants were required to identify the false information tweet against a similar lure tweet. As expected, participants who had not seen the false information (F-R-) performed at chance levels on this test (Table 6.4). By contrast, participants who were exposed to the false information (F+R+, F+R-), performed almost flawlessly. This was even the case for the retraction group, highlighting that the retraction had not overwritten the false information in the memory store.

6.1.3 Discussion

In previous chapters, we have examined whether information stored in memory might be updated via an overwriting mechanism. However, in the present study, we turned our attention to other factors that might contribute to the updating of erroneous knowledge. Specifically, we sought to evaluate whether the continued influence effect (CIE; Seifert, 2002) could be eliminated when a retraction message was accompanied by an alternative account that explained more features of the event compared to the account associated with the false information. In order ot achieve this, we ensured that all four auxiliary details presented in the narrative were more consistent with the alternative theme than with the false theme. According to the causal coverage hypothesis, this should prevent participants from relying on false information after it has been retracted because a plausible and coherent replacement account is readily available. However, despite enhancing the causal coverage of the alternative theme, the CIE was not eliminated in this experiment.

The alternative themes we provided were clearly plausible. In the baseline conditions (F-R-) participants made a high number of alternative theme references, and as expected, rarely made spontaneous references to the false theme (Figure 6.4, Figure 6.5). For the group who also received false information without a retraction, reliance of the alternative theme was partly reduced and the number of references to false information increased (Manipulation Check; F+R- > F-R-). Therefore, we were able to establish the necessary conditions under which a CIE could emerge. As demonstrated previously (e.g., Ecker et al., 2011b; H. M. Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988), the retraction message reduced reliance on the false information (Hypothesis 1; F+R+ < F+R-). We also observed a small recovery in reliance on the alternative theme. However, the retraction message did not eliminate the influence of the false information: the number of false theme references was higher in the group who experienced false information with a retraction than in the group who never experienced the false information (Hypothesis 2; F+R+ > F-R-). Therefore, like others (Ecker et al., 2010, 2011a; H. M. Johnson & Seifert, 1994; Wilkes & Reynolds, 1999), we found that making an alternative account readily available was capable of attenuating, but not eliminating the CIE.

Although the causal coverage manipulation did not eliminate the CIE, it could still have enhanced the efficacy of the retraction. To evaluate this, an additional experiment should be run in which the false theme has greater causal coverage than the alternative theme. Under these conditions, one would expect the CIE effect to be more pronounced because the most plausible and coherent account of the event is based on the false information, thus introducing greater resistance to the retraction. Nevertheless, the fact that participants continue to rely on false information, even when it has been retracted and a plausible, coherent alternative is readily available, suggests that other important factors also contribute to the CIE. One candidate is a failure of 'strategic memory processing' (Johnson, Hashtroudi, & Lindsay, 1993; Roediger & McDermott, 1995). When individuals recall the false information, they may fail to adequately search memory for other relevant information, such as the retraction message. Although data from CIE studies (and the present experiment) show that participants can in principle recall the retraction message, this does not necessarily mean that they access this 'negation tag' whilst retrieving the false information (Ecker et al., 2011b). This account could explain why warning participants that they will encounter false information can attenuate the CIE (Ecker et al., 2010): it may encourage participants to adopt more rigorous strategic monitoring when answering the inference questions. Interestingly, in one study the dual use of a prior warning and an alternative message still did not eliminate the CIE (Ecker et al., 2010). However, as discussed in the introduction, this could be because the alternative account had low causal coverage relative the false theme. Future investigations could examine the impact of a dual manipulation involving a prior warning and an alternative account with high casual coverage.

Despite failing to eliminate the CIE, we found that the retraction did partly reduce reliance on the false information, as others have shown repeatedly (e.g., H. M. Johnson & Seifert, 1994; Ecker et al., 2011b; cf. Wilkes & Leatherbarrow, 1988). We sought to establish whether this memory updating effect entailed any storage-based impairment (Hypothesis 3; F+R+ > F-R- and F+R+ = F+R-), as implied by a trace overwriting account (Dudai, 2009; Loftus, 1979a; Wilkes & Leatherbarrow, 1988). As expected, the F-R- group, who had not seen either the false information tweet or the lure tweet, performed at chance levels on a forced-choice recognition test (see Chapter 4; M. McCloskey & Zaragoza, 1985a). By contrast, performance in the two groups who had seen the false information tweet was equivalent and almost flawless (F+R- and F+R+). This evidence clearly demonstrates that the retraction message had not caused any storage-based memory impairment, and stands in stark contrast to destructive updating, editing, or overwriting accounts (Dudai, 2009; Loftus, 1979a; Wilkes & Leatherbarrow, 1988).

An alternative explanation is that participants were able to voluntarily suppress the to-be-forgotten false information through a process of intentional forgetting (M. C. Anderson & Neely, 1996; R. A. Bjork, 1989; MacLeod, 1998). In an illustrative intentional forgetting paradigm (also known as 'directed forgetting'), participants learn a list of items (e.g., *hand*, *duck*, *horse*, etc.). After each item, they are instructed to either 'remember' or 'forget'[45]. Subsequently, participants complete a recall test for all of the items they have seen. The typical finding is that recall of to-be-forgotten (TBF) items is lower than for to-be-remembered (TBR) items. An early 'erasure' hypothesis, posited that individuals had deleted some of the TBF items from memory (Muther, 1965). However, this account was quickly undermined by several studies showing that TBF materials are available in the memory store, as evidenced by intact performance on recognition tests (e.g., Block, 1971; Geiselman, Bjork, & Fishman, 1983). Therefore, individuals who accept and implement the intentions of the retraction message in the CIE paradigm may be able to voluntarily suppress (but not overwrite) the false information, thus avoiding its usage on the inference questions, but allowing its presence to be revealed on a recognition test. Reducing the CIE might therefore require intentional forgetting of the false information, but it is unclear why some participants appear to be more successful at this than others.

Finally, an intriguing finding that warrants further discussion is that participants often made references to *both* the alternative theme and the false theme. This was most evident in the F+R- condition where participants had no reason to disregard any of the information they had read. A close examination of participants' responses revealed many examples where false

[45] Alternatively, in the *list-method*, one group of participants is instructed to forget the first half of a list of items and remember the second half, whereas a second group are instructed to remember both parts of the list (MacLeod, 1998).

and alternative themes were referred to alongside each other (e.g., "The cage was left open and there was a hole in the perimeter of the fence"). Some participants displayed a tendency to generate novel 'blended' explanations, integrating information from both themes (e.g., "I think that they will find out that the man with the 'animal rights' shirt convinced the work experience lads that they should free the animals"). This is consistent with evidence that, rather than revising their beliefs, individuals often prefer to generate new explanations that resolve inconsistencies within their mental models (Johnson-Laird, 2012). The false and alternative themes appear to be lend themselves to this process because they are not mutually exclusive. It is not impossible, as one participant suggested, that the warehouse fire was caused by an incompetent worker leaving gas cylinders (false theme) next to an unattended stove in the staff kitchens (alternative theme). This raises a potential new hypothesis: how would participants handle the two themes when one rendered the other impossible? For example, returning to the scenario outlined earlier involving a plane crash and a drunken pilot: would the alternative theme eliminate reliance on the false information if this was in fact the maiden voyage of a new 'pilot-less' jet that suffered a catastrophic computer failure? Presumably this would force participants to choose between the two accounts, rather than attempting to blend them. Finally on this issue, it is important to note that although some theorists have proposed that incorporating new information into memory can lead to the creation of "blended memory representations" (e.g., Loftus et al., 1985), the present data offer no reason to assume that the blending process involves physical amalgamation of traces (also see M. McCloskey & Zaragoza, 1985b). These findings could just as well reflect some 'online' retrieval-based synthesis of information that is currently active in memory (Johnson-Laird, 2012).

In summary, the CIE suggests that erroneous information stored in memory is not routinely updated via an overwriting process. In the present study, we examined whether the CIE could be eliminated via an alternative means: the provision of a readily available alternative account that provides superior causal coverage to the false information account. Our findings indicate that these conditions are not sufficient to abolish the effect, suggesting that the lack of an adequate alternative explanation is not the only factor that determines the continued influence of false information.

Chapter 7

General Discussion

> "And in it all, where did the truth end and error begin?"
>
> — Verne, *Journey to the Centre of the Earth*

For over a century, scientists have sought to establish how the memory system maintains knowledge representations over time (*persistence*) whilst simultaneously retaining a capacity for updating knowledge when it is outdated or erroneous (*plasticity*). In this thesis, I have examined the provocative idea that memory traces can be overwritten with new information, especially during transient periods of retrieval-induced plasticity that occur when a trace undergoes *reconsolidation* (Nader & Hardt, 2009). If reconsolidation enables overwriting of memory traces in humans, the clinical (Schwabe et al., 2014), ethical (Hui & Fisher, 2014), and theoretical (Nader et al., 2000b) implications could be profound. Nevertheless, after evaluating the existing evidentiary landscape (Chapter 2), and subjecting the theory to several strong tests in our own experiments (Chapters 3, 4, and 5), it is necessary to conclude that the extant evidence for reconsolidation-mediated memory updating is remarkably tenuous. In this final chapter, we will summarise the main findings of the thesis and provide an integrative discussion of key issues.

7.1 Summary and discussion of findings

In Chapter 1, we first highlighted the inherent complexity of empirical investigations into memory updating and forgetting. Three stages that intervene between learning and performance were introduced: (re)storage, retrieval, and conversion (section 1.1). Our framework stressed the important distinction between a latent *memory trace* (i.e., the physical representation of encoded information) and *ecphoric memory* (i.e., the state of activity generated by the synergistic interaction of retrieval cues and multiple reactivated memory traces; Moscovitch, 2007; Semon, 1921; Tulving, 1983b). The efforts of memory researchers to unweave this complex tapestry are constrained by the limitations of the technologies available to them. For example, it is currently not possible to directly verify that a memory trace has been modified or erased (although for promising recent developments see Josselyn et al., 2015), and researchers must largely rely on indirect inferences based on behavioural measures (M. J. Watkins, 1990). This issue has surfaced repeatedly in the enduring search for the loci of forgetting (Gold & King, 1974; Gold, 2006; R. R. Miller & Springer, 1973; R. R. Miller & Matzel, 2006) and remains problematic (Nader, 2006; Sara & Hars, 2006; Squire, 2006).

In early studies of retroactive interference, it was found that learning new information could adversely influence retrieval of previously learned 'old' information on a subsequent memory test (Crowder, 1976; Postman & Underwood, 1973). This was initially demonstrated in a series of experiments conducted by Müller and Pilzecker (1900) who also observed that the disruptive effects of new learning appeared to diminish as the post-acquisition time-interval increased. This temporal gradient was taken to indicate the status of a progressive stabilisation process, 'consolidation', necessary for the fixation of the memory trace. Subsequent investigations in non-human animals used invasive pharmacological and electrophysiological interventions to target the putative physiological substrates of consolidation. These studies largely confirmed the existence of a temporal gradient across multiple species, paradigms, and intervention types (Glickman, 1961; McGaugh, 1966). Later refinements of consolidation theory adopted a dual trace model (Hebb, 1949; McGaugh & Dawson, 1971), in which a short-term memory (STM) trace consisted of initial reverberatory activity in local neural circuits, and was followed by protein-synthesis dependent structural changes at the synaptic level (long-term memory, LTM). Critically, it was assumed that consolidation was a uni-directional process. Once the intervention lost its effectiveness, the trace had by definition 'consolidated' and was thought to be immune to any subsequent attempts at interference.

Consolidation theory has persevered through more than a century of empirical examination, but it is unclear if it has prevailed. A number of studies conducted during the 1960s and 1970s found that reactivating a previously consolidated memory trace could apparently render the trace vulnerable to disruption once again (e.g., Misanin et al., 1968). These cue-dependent amnesia (CDA) effects represented a serious challenge to the assumption that amnesic agents were only effective during a brief post-acquisition 'consolidation window' (Lewis, 1976; R. R. Miller & Springer, 1973; Riccio & Richardson, 1984). However, some research teams had trouble replicating CDA effects (e.g., Dawson & McGaugh, 1969) and alternatives to consolidation theory failed to gain traction (Lewis, 1979). Nevertheless, a more recent demonstration of CDA (Nader et al., 2000a) received considerable attention in the neuroscience community, and the debates of the 1960s and 1970s rapidly resurfaced (see Riccio et al., 2006; Sara & Hars, 2006). Critically, this new report of CDA was accommodated within the consolidation framework, rather than seen as a direct challenge to it (Dudai, 2000; Nader et al., 2000b; Sara, 2000b). *Re*consolidation theory suggests that reactivation can destablise a previously consolidated trace, reinstating a state of plasticity until the trace restabilises ('reconsolidates'). The years following the study of Nader et al. saw reconsolidation theory draw widespread empirical support from across species, experimental paradigms, and intervention types (for review see Besnard et al., 2012; Nader & Hardt, 2009; Tronson & Taylor, 2007).

However, in several respects, reconsolidation theory is undermined by the same issues that posed difficulties for the consolidation account (Chapters 1 and 2). For example, the consistent finding that trace-dependent performance is intact on tests of STM, but not on tests of LTM, can be reasonably accommodated in the context of protein-synthesis inhibitor (PSI) studies (see Davis & Squire, 1984), but seems more problematic in studies administering electroconvulsive shock (ECS) studies (see subsubsection 1.2.7.6; R. R. Miller & Springer, 1973). PSIs are thought to prevent (re)stabilisation of the LTM trace directly, but have no effect on the STM trace. ECS on the other hand eliminates the STM trace, which prevents subsequent (re)stabilisation of the LTM trace (R. R. Miller & Springer, 1973; R. R. Miller & Matzel, 2006). In consolidation studies, the finding of intact STM after only brief acquisition-intervention delays (e.g., 10s) seems to indicate that the formation of a structural memory representation occurs extremely rapidly (R. R. Miller & Springer, 1973).

The (re)consolidation account suggests that the disruption caused by post-acquisition and post-reactivation interventions is permanent, because the target trace had not been allowed to (re)stabilise, and is consequently impaired, or even erased (Nader, 2003a). However, many studies have found that experimentally-induced amnesia can be partially alleviated, either spontaneously with the passage of time, or via specific 'reminder treatments' such as reinstatement or renewal (see subsubsection 1.2.7.8; Spear, 1973; R. R. Miller & Springer, 1973; Riccio & Richardson, 1984). This echoes findings from extinction and counter-conditioning studies which have repeatedly found that superficially compelling cases of 'unlearning' were not due to unlearning at all, but instead reflected retrieval-based forgetting of the target CS-US association (see subsection 1.2.3; Bouton, 2002). Recovery effects have been reported on numerous occasions and appear to provide compelling evidence against a storage-based mechanism of memory impairment. However, they have not always proved reliable in subsequent replication attempts (e.g., Duvarci & Nader, 2004) suggesting that the conditions under which they arise are poorly understood. Moreover, the degree to which they actually pose a serious problem for (re)consolidation theory has been challenged by some theorists (Gold & King, 1974; Nader & Wang, 2006; Squire, 2006).

One of the most exciting and controversial aspects of reconsolidation theory is the idea that it represents a memory updating mechanism (Dudai, 2009; Exton-McGuinness et al., 2015; J. L. Lee, 2009). Specifically, it has been proposed that transient 'reconsolidation windows' provide an opportunity to modify or 'overwrite' the contents of memory traces, allowing routine maintenance of their accuracy and relevance in a changing environment (J. L. Lee, 2009). In a 2010 article, Hardt, Einarsson, and Nader outlined how this aspect of reconsolidation could act as a bridge between the cognitive and neurobiological approaches to memory research. They suggested that the field of cognitive psychology is replete with demonstrations that 'memories' are 'malleable' entities citing, for instance, retroactive interference effects in the A-B, A-C paired-associates (Postman & Underwood, 1973) and misinformation paradigms (Loftus, 1979a). They commented on how these findings are at odds with consolidation theory, which suggests that memory traces, once initially stabilised, remain static and fixed. Reconsolidation theory, on the other hand, imbues memory traces with a capacity for plasticity, which can explain how they might be updated by new learning (Hardt et al., 2010).

The problem here is that there is compelling evidence in both cases that no modification of memory traces has occurred. In the A-B, A-C paired-associates paradigm for example (for details see Chapter 1), the interference effect can be mitigated by asking participants to provide both responses (e.g., J. M. Barnes & Underwood, 1959), diminishes over time (e.g., Wheeler, 1995), and can even be eliminated when a recognition test is employed (e.g., R. C. Anderson & Watts, 1971). Similarly, in the misinformation paradigm (for details see Chapter 4), the interference effect can also diminish over time (Chandler, 1991), and be eliminated (M. McCloskey & Zaragoza, 1985a) when appropriately designed recognition tests are used. In other words, these retroactive interference effects seem to be best characterised as retrieval-based rather than storage-based impairments, and neither phenomenon requires that the memory trace itself be a malleable entity (Chandler & Fisher, 1996; Crowder, 1976). More broadly, it is well known that the inaccessibility of a memory trace (i.e., retrieval-failure) does not necessarily reflect the unavailability of the trace (i.e., storage-failure). Many instances of forgetting appear to be cue-dependent: that is, they can be alleviated through provision of effective retrieval cues (see subsection 1.2.4; Tulving & Pearlstone, 1966; Tulving, 1974). This is also readily apparent from the extensive empirical investigations into extinction and counter-conditioning (mainly in non-human animals, see subsection 1.2.3; Bouton, 2002; Bouton & Moody, 2004). The recovery effects obtained in these studies indicate that it is unnecessary to 'overwrite' or 'unlearn' an existing memory association in order to make way for an association representing a conflicting behavioural response. Overall, there is considerable evidence to suggest that when new learning leads to forgetting, or memory updating, it does not require modification of existing memory traces.

How do studies fare when they directly test the theory of reconsolidation-mediated memory updating? In Chapter 2, we addressed this by conducting a comprehensive review of all published investigations of human reconsolidation that have employed behavioural interventions. After defining ten key criteria necessary for a compelling demonstration of reconsolidation, we systematically evaluated the extent to which these criteria were met in the current evidentiary landscape. We also conducted a more in-depth examination of several case studies (also see Chapters 3 and 4). Our findings indicate that studies rarely meet more than a few of the critical criteria, and no study met more than six. As such, we concluded that there are presently no compelling demonstrations of reconsolidation-mediated memory updating in human participants.

Our review also took a 'meta-theoretical' perspective, and examined how the field has handled instances of theoretical success and failure (section 2.3). We suggested that frequent appeals to unverified 'boundary conditions' has stalled rather than facilitated theoretical progress because of a lack of systematic empirical verification. An immediate priority for the field will be to establish a protocol in which reconsolidation effects can be reliably observed. An effect can be considered relatively reliable when it has been estimated with high precision and can be obtained in independent high-powered direct replication attempts. To achieve these aims, the field will need to adjust many of its informal conventions. For example, precise effect size estimates will likely require larger sample sizes than are typical (see Figure 2.6) and direct replication will only be possible if experimental materials and software programs are made openly available for other researchers to use. The focus on repeating previous studies may seem less exciting than the pursuit of more novel ventures, but replication is a vital ingredient of scientific discovery because it separates signal from noise (Ioannidis, 2014). Ultimately, systematic replication efforts will be the only way to identify genuine boundary conditions and avoid the considerable theoretical confusion that results from the widespread proliferation of *ad hoc* hypotheses.

Chapter 3 provides an example of this approach in a series of experiments intended to replicate a prominent finding interpreted as evidence for reconsolidation-mediated memory updating in humans (Walker et al., 2003). This particular procedural learning study met six of the reconsolidation criteria outlined in our review, and was hailed as a 'landmark discovery' in the reconsolidation community (Nader, 2003b, p. 572). However, across four direct replications (Experiments 1-4, $N = 64$) and three conceptual replications involving declarative recall (Experiments 5-7, $N = 48$), we did not observe the critical impairment effects obtained in the original study, and as such, found no evidence to support the claim that reconsolidation enables human memory updating.

In Chapter 4, we applied a similar 'replication battery' approach to a case reporting reconsolidation-mediated updating in a misinformation paradigm (J. C. K. Chan & LaPaglia, 2013). In their study (specifically $\text{Experiment}_{\text{CL}}$ 6), Chan and LaPaglia adapted the traditional misinformation paradigm in order to meet various reconsolidation criteria (5 were met). Specifically, after 'witnessing an event' (watching a video) on Day 1, participants were given a reminder test on Day 2, immediately prior to listening to an audio narrative describing the Day 1 video. The narrative referred to some critical items in a misleading way ('misled' condition) and referred to other items in a neutral way ('control' condition). Finally, participants were tested on Day 3 using a true/false recognition test. Under these conditions, the researchers reported that performance was significantly higher for neutral items compared to misled items, i.e., a 'misinformation effect' (Loftus, 1979a) was observed. By contrast, in a group of participants who did not complete the reminder test on Day 2, no significant difference between neutral and misled items was observed. It was concluded that reconsolidation-mediated overwriting of the memory trace representing the Day 1 video had occurred in the reminder group, but not in the no reminder group.

There are at least two important reasons to be skeptical of this conclusion. Firstly, the critical reminder by intervention interaction was not tested (see Chapter 2, Criterion 7), and secondly, the classic misinformation literature indicates that the effect can be eliminated when an appropriately designed recognition test is employed. Specifically, although misinformation effects are readily obtained using a forced-choice recognition test that includes the event item and the misled item as options (the "Original Test", Loftus et al., 1978), they can be eliminated when the misled item is excluded as an option, and there is instead a forced-choice between the event item and a novel foil item (the "Modified Test", M. McCloskey & Zaragoza, 1985a). The 'True/False' recognition test used by Chan and LaPaglia was similar to the Modified Test but was susceptible to influence by factors unrelated to memory impairment, and therefore did not provide an optimal test of reconsolidation theory (see Chapter 4 for details).

In Experiment 8, we first established that the misinformation effect could be obtained in a hybrid reconsolidation-misinformation paradigm when the Original Test was employed. We asked a second group of participants to respond on the Modified Test, and found that the effect was eliminated, replicating earlier findings obtained under non-reconsolidation conditions (e.g., M. McCloskey & Zaragoza, 1985a). In Experiment 9, we asked all participants to respond on both the True/False and Modified Tests. We found a small misinformation effect on the True/False Test, partially replicating the findings of Chan and LaPaglia (2013). However, as in Experiment 8, we found that the misinformation effect was eliminated on the Modified Test. In other words, participants were perfectly capable of recognising the event item against a novel lure, indicating that, contrary to reconsolidation theory, the trace representing the event item was intact. It would seem that the type of test used at retrieval can moderate the presence or absence of impairment, consistent with the idea that the misinformation effect represents a retrieval-based rather than storage-based phenomenon (Chandler & Fisher, 1996; Zaragoza et al., 2006). Finally, in Experiment 10, we replicated the Modified Test outcomes obtained in Experiments 8 and 9, and also found that the nature of the reminder stage (no reminder, retrieval practice, or ‘subtle’ reminder) had no appreciable impact on final test performance. Overall, the findings of these three experiments suggest that misinformation effects are not caused by reconsolidation-mediated memory updating.

One outstanding issue in this line of research concerns the broader field of misinformation studies that have used the Modified Test outside the context of reconsolidation theory. Although the majority of Modified Test outcomes do not indicate a reliable difference between misled and control conditions (as in Experiments 8-10), a small number of studies have found misinformation effects. In fact, an early meta-analysis suggested that when a large number of Modified Test outcomes were considered together, a small but statistically significant misinformation effect could be detected (Payne et al., 1994). Therefore, it could be that geunine impairment effects occur, but it is unclear exactly what circumstances are necessary (also note that the Modified Test cannot differentiate between a retrieval- or storage-based impairment effect). Moreover, the meta-analytic approach employed was problematic because it did not account for the precision of individual studies. Unfortunately, we were unable to conduct additional analyses of our own because of an absence of detailed statistical reporting in the original sources.

To further complicate matters, some researchers have suggested that the Modified Test may not be sufficiently sensitive to detect small impairment effects (Belli et al., 1992; Loftus et al., 1985). A difference between control and misled conditions detected using the Modified Test is likely to be 50% smaller than any actual impairment effect because half of the responses made for impaired items are likely to be correct guesses (Belli et al., 1992; M. McCloskey & Zaragoza, 1985a). Therefore, one argument against the conclusions of Chapter 4 is that the experiments did not have sufficient statistical power to detect small effects.

There are counter-arguments to this proposal, however. For example, studies included in the meta-analysis that had larger sample sizes tended *not* to find misinformation effects (see Figure 4.4). Furthermore, a meta-analytic Bayes Factor applied to Modified Test outcomes in Experiments 8-10 indicated reasonably strong evidence in favour of the null hypothesis relative to the reconsolidation hypothesis (H_0 was approximately 12 times more likely than H_1 given the data). Nevertheless, there is a degree of ambiguity about the extant evidence that warrants further investigation. We have proposed that the field pools its resources and expertise in a pre-registered, multi-lab, Registered Replication Report (Simons et al., 2014) in order to identify the degree to which misinformation effects can be reliably detected using the Modified Test.

In Experiment 11 (Chapter 5), we examined forward/backward testing accounts and the reconsolidation theory in the context of updating knowledge about semantic facts. Specifically, we reasoned that if reconsolidation-mediated updating is a genuine phenomenon, then prior retrieval of naturally occurring semantic misconceptions should facilitate their correction with accurate information via a process of 'overwriting'. Conversely, a substantial body of evidence indicates that retrieval of target information can actually increase the likelihood of its recall on a subsequent memory test (Roediger & Butler, 2011), potentially reinforcing the erroneous knowledge. In addition, it has been reported that retrieval practice can enhance learning of subsequently presented information (Pastötter & Bäuml, 2014), potentially facilitating acquisition of the accurate information. As such, contrasting theoretical predictions were available as to whether prior retrieval would help or hinder the correction of semantic misconceptions.

We firstly elicited naturally occurring semantic misconceptions by asking participants to respond to a large battery of 'quiz' style questions, for example, 'What is the name of a dried grape?' (answer: 'Raisin'). On Day 2, participants were asked to answer some of these questions again (i.e., retrieval practice), presented with questions along with their previous responses (i.e., restudy), or simply rested for a short period (i.e., no reminder), immediately prior to receiving corrective feedback. Finally, on Day 3, participants were asked to respond to the questions again in three phases. Firstly (dominance phase), in order to examine knowledge updating we asked them to provide what they considered to be the correct answer to the question. Secondly (persistence-new phase), to confirm that they had learned the corrective feedback, we specifically asked them to provide the information we had given them on Day 2. Finally (persistence-old phase), in order to establish whether their previous erroneous knowledge had been forgotten, we asked them to specifically provide the response they had given us on Day 1.

Participants were highly successful at updating their knowledge, responding mostly with the accurate information during the dominance and persistence-new phases. Contrary to both reconsolidation and forward-testing accounts, performance during these phases was not impacted appreciably by the reminder manipulation. In other words, prior retrieval of erroneous knowledge was not necessary for, nor did it facilitate, updating with accurate information. In the critical persistence-old phase, we examined whether updating following a reminder would result in the overwriting (reconsolidation account) or enhancement (testing account) of the erroneous knowledge. The findings clearly demonstrated that most old knowledge had been retained. Contrary to reconsolidation theory, recall of old knowledge was higher in the reminder (restudy, retrieval practice) conditions relative to the no reminder (rest) condition. The pattern of results across the three phases was maintained on a three week follow-up test indicating that both corrective feedback and previous misconceptions had largely been retained. Interestingly, the dominance of the corrective feedback had partly subsided and responding with misconceptions had increased by a small magnitude.

Overall, these findings provide further evidence against the reconsolidation account. Firstly, memory updating does not appear to be restricted to privileged ‘windows of plasticity’ whilst traces are undergoing reconsolidation. Secondly, memory retrieval, in the form of testing or restudy, can enhance the target information rather than rendering it vulnerable to interference. Thirdly, memory updating does not require impairment or erasure of competing information that already exists in the memory store. The findings of the three-week follow-up test also provide initial evidence that abandoned knowledge can linger in the memory store and start to recover its dominance over time. This raises the possibility that even after successfully learning corrective feedback, erroneous information could persist, and continue to influence an individual's judgments and decision-making.

Nevertheless, the findings of this single experiment should be considered relatively exploratory and await confirmation in a pre-registered, high power replication. Moreover, it will be necessary to address the possible ceiling effects that could have masked a performance advantage for the retrieval practice condition relative to the restudy condition. The intriguing observation that semantic misconceptions started to regain their dominance after a 3-week period also warrants further scrutiny. An obvious manipulation would be to evaluate whether this apparent regression to previous knowledge is exacerbated at longer time intervals. It will also be interesting to evaluate whether abandoned, but lingering erroneous knowledge can have other more subtle consequences. For example, if participants were asked to write an essay about the target topic rather than directly answer a question about it, this might reveal a type of continued influence effect akin to the one observed in Experiment 12.

In Chapter 6, we turned our attention to the continued influence effect (CIE): the phenomenon whereby individuals continue to rely on false information even after they have been explicitly told to discount it (Seifert, 2002). CIEs appear to demonstrate quite clearly that information stored in memory cannot simply be overwritten even when it is clear that the information is erroneous and requires correction. The societal repercussions of this phenomenon are potentially very serious (Lewandowsky et al., 2012b). Consider, for example, that a jury considering its verdict might continue to rely on false information even when a judge has specifically instructed them to disregard it (Fein et al., 1997). A typical laboratory demonstration of CIE involves participants reading a media report (e.g., regarding a fire at a warehouse) during which they are exposed to some false information (e.g., 'the storeroom contained paint cans and gas cylinders'), which is later explicitly retracted (e.g., 'the storeroom did not contain paint cans and gas cylinders, it was empty'). Despite acknowledging the retraction message, participants still tend to make references to the false information when responding to questions about the event (Seifert, 2002). Typically, the retraction is at least partly effective, it is just not sufficient to eliminate the influence of the false information. As such, CIEs appear to demonstrate that information stored in memory cannot simply be erased even when it is clear that the information is erroneous.

The CIE has proved extremely difficult to eliminate entirely (e.g., Ecker et al., 2010). However, one of the most effective methods for mitigating the influence of the false information is to provide an alternative explanatory account of the event (e.g., the warehouse fire was started by an arsonist, H. M. Johnson & Seifert, 1994). In Experiment 12, we examined whether the *causal coverage* of an alternative account would increase its effectiveness, and potentially eliminate the CIE. Specifically, we noticed that in previous studies, the false information tends to be consistent with other details provided about the event. For example, in the warehouse fire narrative, the presence of paints and gas cylinders (the false information) is consistent with auxiliary details such as “oily smoke and sheets of flame”, “intense heat”, “explosions”, and “toxic fumes”. Therefore, even if the false information is retracted, it may still provide the most compelling account of the event, and therefore prove difficult to ignore. We hypothesised that providing an alternative account with superior causal coverage (i.e., explaining more event details relative to the false information account) would eliminate the CIE.

Our findings replicated the CIE: participants who received a retraction message made fewer references to false information than a no-retraction group, but made more references than a control group who had not been exposed to the false information at all. In other words, our attempt to eliminate the CIE by providing an alternative account with high causal coverage was not successful. Nevertheless, additional experiments will be required to ascertain whether the high causal coverage account was more effective than an account with lower causal coverage. In Experiment 12, we also used a ‘modified’ recognition test with the false information message and a novel lure as options (M. McCloskey & Zaragoza, 1985a) in order to examine whether the effectiveness of the retraction could be explained by a trace overwriting account. Our participants displayed near-perfect performance on this test, suggesting that no destructive updating had occurred.

7.2 Adaptive updating

Why is speculation about reconsolidation-mediated memory updating so widespread when there appears to be such a remarkable shortfall of empirical evidence (see Chapter 2)? The answer may lie in an unwarranted conceptual conflation of behavioural and physiological plasticity. It seems self-evident that we can learn to change our behaviours in response to the changing demands of our local environment. An organism that does not update its knowledge may lose access to vital resources such as food, water, or shelter; or, for homo sapiens of more recent times, their bank account. However, it is not necessarily the case that this adaptive *behavioural* plasticity relies directly on *physiological* plasticity (cf. R. A. Bjork, 1978; Dudai, 2009; Kraemer & Golding, 1997). In other words, just because a previously learned behavioural response is 'replaced' by an alternative behavioural response does not *require* modification of memory traces. Nevertheless, there seems to be a widespread assumption that this is exactly how behavioural plasticity is achieved: the world around us is dynamic, and therefore memory traces must also be dynamic if the organism is to be adaptive. From this perspective, the concept of reconsolidation is appealing because the memory trace, previously seen as 'fixed' and 'static' can now be imbued with 'malleability' and plasticity (Dudai, 2009; Hardt et al., 2010; J. L. Lee, 2009)

But there is good reason to reject this conflation and maintain a separation between behavioural and physiological plasticity. A system in which memory content is overwritten when contradictory information is encountered in the environment would arguably be disastrous. The system would be calibrated too far in favour of plasticity at the expense of persistence (i.e., susceptible to catastrophic interference), and would be overly sensitive to temporary fluctuations ('noise') at the expense of global patterns ('signals') that emerge over time. Anderson, Lepper, and Ross (1980, p. 1046) expressed this view clearly: "To be buffeted about by every random piece of disconfirming data or every challenge to the evidential basis for one's beliefs, whether in the course of scientific inquiry or in our daily lives, will frequently prove less adaptive then a tendency to persist in theories that have proven effective over time." Therefore, in principle, it is hard to see how it a system that relies on overwriting to update knowledge could actually operate effectively in the natural world.

Far more useful would be a system that maximises the persistence of traces, in order to preserve a rich repertoire of behavioural responses that can be deployed flexibly depending on the retrieval cues available in the current situation. The system can still achieve behavioural plasticity by continually accumulating new traces. As documented extensively in Chapter 1, there is considerable evidence indicating that this is precisely how the memory system operates (see 'occasion setting', 'response competition', 'retrieval competition', and 'memory-as-discrimination'). For instance, in the counter-conditioning paradigm, an animal might first learn that a cue (e.g., tone) signals an aversive outcome (e.g., shock), and then subsequently learn that the same cue signals an appetitive outcome (e.g., food). Rather than overwriting information, the animal appears to be perfectly capable of simultaneously maintaining both associations, despite the fact that they elicit opposing behavioural responses (fear response or appetitive response). Moreover, by manipulating the availability of retrieval cues (e.g., physical context), the researcher can elicit either response from the animal (Bouton & Moody, 2004).

7.3 Closing remarks

The situations in which knowledge updating is required are commonplace. Doctors need to continuously adjust the treatments they prescribe on the basis of new evidence, jurors need to disregard biased testimony when instructed to do so by a judge, and scientists need to revise their theories in light of new discoveries. There can be little doubt that the memory system has the *functional* capability to update. The controversial question is how it achieves this aim.

In the 1970s it was proposed that exposure to contradictory information could lead to the overwriting or 'destructive updating' of existing memory traces (Loftus, 1979b), and prior to that, the similar concept of 'unlearning' had gained considerable theoretical attention (Melton & Irwin, 1940; Rescorla & Wagner, 1972). Lately the embers of these old ideas have been rekindled. Trailing the meteoric rise of reconsolidation theory, there has been widespread speculation that memory traces become amenable to modification during transient periods of retrieval-induced plasticity (e.g., J. L. Lee, 2009; Exton-McGuinness et al., 2015).

However, as key criteria that ensure a robust test of the theory are rarely met in empirical investigations, the evidentiary basis for this claim is far from compelling. Moreover, the widespread use of *ad hoc* hypotheses that reduce explanatory coverage, and the absence of systematic replication efforts needed to verify the reliability of published evidence, has merely led to theoretical obfuscation rather than theoretical progress.

Throughout this thesis I have made several appeals to the principles of open and cumulative science, specifically: open data, pre-registration, and direct replication. These are not tangential issues; it is quite clear that their neglect has had tangible consequences for the field. Openness is the pre-requisite of a scientific system that self-corrects and without it, how will we ever know where truth ends and error begins?

Appendix A

Precise operationalisation of dependent variables (Experiments 1-4)

It should be noted that the precise operationalisation of the dependent measures reported in the original study was ambiguous: Performance measures were the number of complete sequences achieved ('speed'), and the number of errors made relative to the number of complete sequences achieved ('accuracy'). The senior author of the original research team confirmed the following definitions. *Speed* was the number of complete sequences achieved during a 30s trial plus any partial sequence the participant was completing when the trial was terminated. For example, a participant who performed 15 complete sequences, and had just entered two correct items when the trial terminated, would receive a speed score of 15.4 $(15 + \frac{2}{5})$. *Accuracy* was $1 - \frac{\text{errors}}{\text{speed}}$, where a single error was defined as any string of up to five contiguous incorrect items that did not match the target sequence. For example, three contiguous incorrect items would constitute a single error, but six contiguous incorrect items would constitute two errors.

Note that, under this scheme, it is technically possible for a participant to incur a negative accuracy score on an individual trial if error exceeds speed. This could substantially bias between-stage comparisons, as accuracy scores should only range between 0 and 1. Across the four experiments reported here, five trials with negative accuracy scores were identified (<0.003% of total trials) and converted to zero. This did not impact the overall pattern of results.

Appendix B

Additional analyses for Chapter 3

B.1 Influence of session times and participant age (Experiments 1-4)

Two minor procedural differences between the original study and the direct replications (participant age range and time of testing) were evaluated to see whether they influenced the findings (see Walker & Stickgold, 2016). Time of testing in the direct replications (rounded to the nearest hour: $Mdn = 15.00$ hours, $SD = 1.833$) differed only slightly from Walker et al. (13.00 hours) and was not significantly correlated with RSs ($r = 0.12$, $p = 0.355$). Therefore, time of testing does not appear to account for the absence of a reconsolidation effect.

Participant age in the direct replications ($Mdn = 22$ years, range = 18–54 years) covered a larger range than Walker et al. (*Mdn* unknown, range, 18–27 years). A reanalysis of RSs for only those participants who fell within the 18–27 age bracket ($n = 48$), showed that there was still no substantial impairment [$M = -2.05$, $SD = 10.51$; $t(47) = -1.35$, $p = 0.092$]. Therefore, participant age does not appear to account for the absence of a reconsolidation effect.

B.2 Impact of counter-balancing (Experiments 1-4)

To establish whether the counterbalancing procedures influenced the RS, we used a series of one-way ANOVAs separately with test order (A, B) or sequence order (X, Y) as a between-subjects factor and RS (separately for accuracy and speed), as a dependent variable. There was no significant main effect of sequence order on RS accuracy, $F(1,62) = 0.004$, $p = 0.948$, or RS speed, $F(1,62) = 0.224$, $p = 0.638$. There was also no significant main effect of test order on RS accuracy, $F(1,62) = 0.655$, $p = 0.421$; however, test order did influence RS speed significantly, $F(1,62) = 5.320$, $p = 0.024$. Follow-up one-sample t tests (two-tailed) indicated that RS speed was significantly *higher* than zero for test order A [$RS = 7.482$, $SD = 13.479$,

$t(35) = 3.331$, $p = 0.002$] and did not differ significantly from zero for test order B [$RS = 0.448$, $SD = 10.044$, $t(27) = 0.236$, $p = 0.815$], confirming that there was no reconsolidation effect in either condition.

B.3 Impact of counter-balancing (Experiments 5-7)

To establish whether the counterbalancing procedures influenced sequence similarity scores at the Test stage, two separate two-way ANOVAs with Experiment (5, 6, 7) and either test order (A, B) or sequence order (X, Y) as between-subjects factors. There was no significant main effect of test order, $F(1{,}42) = 0.466$, $p = 0.498$, or interaction between test order and experiment $F(2{,}42) = 0.723$, $p = 0.491$, and no main effect of sequence order $F(1{,}42) = 0.714$, $p = 0.403$ or interaction between sequence order and experiment $F(2{,}42) = 0.162$, $p = 0.851$.

Appendix C

Table C.1: Critical items for Experiments 8-10.

Critical Item	Question	Event item		Post-event item		Novel item
		Video version		Narrative version		
		V1	V2	N1/N3	N2/N4	
1	What was the name of Eric's company?	RJ's electricians	AJ's electricians	Misleading	Neutral	KJ's electricians
2	In what state was the bed in the first bedroom?	made	unmade	Neutral	Misleading	stripped
3	What did Eric drink a can of?	Coke	Pepsi	Misleading	Neutral	Fanta
4	What colour was the mug next to Eric when he rummaged through papers?	white	yellow	Neutral	Misleading	blue
5	In the second bedroom, what colour was the cap that Eric tried on?	black	blue	Misleading	Neutral	red
6	What magazine did Eric read?	Time	Newsweek	Neutral	Misleading	The Economist
7	How did Eric check the time?	clock	watch	Misleading	Neutral	computer
8	In the lounge which tower was on the picture Eric looked at?	Leaning Tower of Pisa	Eiffel Tower	Neutral	Misleading	Tower of London

Appendix D

Table D.1: Filler items for Experiments 8-10.

Filler item	Question	Event and post-event item	Novel item
		All versions of video and narrative	
1	In the second bedroom, what did Eric test?	power point	light fitting
2	In which room did Eric read the note from the homeowner?	hallway	kitchen
3	What colour was Eric's van?	blue	red
4	What did Eric eat?	apple	banana
5	What did Eric find the house key under?	flower pot	door mat
6	What did Eric look through in the lounge?	photo album	journal
7	What did Eric play?	CD	video
8	What did Eric steal from the bathroom?	pills	perfume
9	What did Eric steal in the second bedroom?	a ring	money
10	What tool did Eric use in the kitchen?	screwdriver	pair of pliers
11	What type of jewelry did Eric steal in the first bedroom?	pair of earrings	necklace
12	What was Eric wearing on the lower half of his body?	jeans	overalls

Appendix E

Table E.1: Counterbalancing conditions for Experiments 8-10.

Condition	Video version	Narrative version
A	V1	N1
B	V1	N2
C	V2	N3
D	V2	N4

Appendix F

Test instructions for Experiments 8-10.

F.1 True/False Test Instructions

“You will now be presented with a series of statements. If you think a statement accurately represents what happened in the video, then choose the ‘TRUE’ option. If you do not think the statement accurately represents what happened in the video, choose the ‘FALSE’ option. After providing a response, please rate your confidence in that response on a scale from 1-100%. Once you have entered a response and made a confidence rating you should press the ‘SUBMIT’ button.”

F.2 Original Test and Modified Test Instructions

“You will now be presented with a series of questions. For each question you will be given two options. Choose the option you think most accurately represents what happened in the video and then rate your confidence in that response on a scale from 1-100%. Once you have entered a response and made a confidence rating you should press the ‘SUBMIT’ button.”

Appendix G

Table G.1: Questions (1-30) and answers for Experiment 11

No.	Question	Correct answer
1	FOR WHICH COUNTRY IS THE YEN THE MONETARY UNIT?	JAPAN
2	IN WHAT ANCIENT CITY WERE THE ‘HANGING GARDENS’ LOCATED?	BABYLON
3	IN WHAT CITY IS THE TALLEST BUILDING IN THE WORLD?	DUBAI
4	IN WHAT EUROPEAN CITY IS THE PARTHENON LOCATED?	ATHENS
5	IN WHAT PROFESSION WAS EMMETT KELLY?	CLOWN
6	IN WHICH CITY IS MICHELANGELO’S STATUE OF DAVID LOCATED?	FLORENCE
7	IN WHICH CITY IS THE UNITED STATES NAVAL ACADEMY LOCATED?	ANNAPOLIS
8	IN WHICH COUNTRY IS ANGEL FALLS LOCATED?	VENEZUELA
9	IN WHICH GAME ARE MEN CROWNED?	CHECKERS
10	IN WHICH TYPE OF SKI RACE DOES THE DOWNHILL SKIER MAKE SHARP TURNS AROUND POLES?	SLALOM
11	OF WHICH COUNTRY IS BAGHDAD THE CAPITAL?	IRAQ
12	OF WHICH COUNTRY IS BUDAPEST THE CAPITAL?	HUNGARY
13	OF WHICH COUNTRY IS BUENOS AIRES THE CAPITAL?	ARGENTINA
14	OF WHICH COUNTRY IS NAIROBI THE CAPITAL?	KENYA
15	OVER WHICH RIVER IS THE GEORGE WASHINGTON BRIDGE?	HUDSON
16	THE DEEPEST PART OF THE OCEAN IS LOCATED AT WHICH TRENCH?	MARIANA
17	THE GENERAL NAMED HANNIBAL WAS FROM WHAT CITY?	CARTHAGE
18	WHAT ANIMAL RUNS THE FASTEST?	CHEETAH
19	WHAT ARE PEOPLE CALLED WHO EXPLORE CAVES?	SPELUNKERS
20	WHAT ARE PEOPLE WHO MAKE MAPS CALLED?	CARTOGRAPHERS
21	WHAT BRAND OF CIGARETTE WAS FIRST TO HAVE THE FLIP-TOP BOX?	MARLBORO
22	WHAT IS THE CAPITAL OF AUSTRALIA?	CANBERRA
23	WHAT IS THE CAPITAL OF CANADA?	OTTAWA
24	WHAT IS THE CAPITAL OF FRANCE?	PARIS
25	WHAT IS THE CAPITAL OF KENTUCKY?	FRANKFORT
26	WHAT IS THE CAPITAL OF NEW YORK?	ALBANY
27	WHAT IS THE CAPITAL OF SWITZERLAND?	BERN
28	WHAT IS THE HIGHEST MOUNTAIN IN SOUTH AMERICA?	ACONCAGUA
29	WHAT IS THE LARGEST PLANET IN THE SOLAR SYSTEM?	JUPITER
30	WHAT IS THE LAST NAME OF THE ASTRONOMER WHO PUBLISHED IN 1543 HIS THEORY THAT THE EARTH REVOLVES AROUND THE SUN?	COPERNICUS

Table G.2: Questions (31-60) and answers for Experiment 11

No.	Question	Correct answer
31	WHAT IS THE LAST NAME OF THE AUTHOR WHO WROTE ‘ROMEO AND JULIET’?	SHAKESPEARE
32	WHAT IS THE LAST NAME OF THE BOY IN THE BOOK ‘TREASURE ISLAND’?	HAWKINS
33	WHAT IS THE LAST NAME OF THE BROTHERS WHO FLEW THE FIRST AIRPLANE AT KITTY HAWK?	WRIGHT
34	WHAT IS THE LAST NAME OF THE COSMONAUT WHO WAS THE FIRST PERSON TO ORBIT AROUND THE EARTH?	GAGARIN
35	WHAT IS THE LAST NAME OF THE CRIMINAL WHO WAS KNOWN AS ‘SCARFACE’?	CAPONE
36	WHAT IS THE LAST NAME OF THE FIRST FLIER TO FLY SOLO AROUND THE WORLD?	POST
37	WHAT IS THE LAST NAME OF THE FIRST PERSON TO COMPLETE A SOLO FLIGHT ACROSS THE ATLANTIC OCEAN?	LINDBERGH
38	WHAT IS THE LAST NAME OF THE FIRST PERSON TO SET FOOT ON THE MOON?	ARMSTRONG
39	WHAT IS THE LAST NAME OF THE FIRST SIGNER OF THE ‘DECLARATION OF INDEPENDENCE’?	HANCOCK
40	WHAT IS THE LAST NAME OF THE MAN WHO BEGAN THE REFORMATION IN GERMANY?	LUTHER
41	WHAT IS THE LAST NAME OF THE MAN WHO INVENTED THE TELEGRAPH?	MORSE
42	WHAT IS THE LAST NAME OF THE MAN WHO SHOWED THAT LIGHTNING IS ELECTRIC?	FRANKLIN
43	WHAT IS THE LAST NAME OF THE MOST POPULAR PIN-UP GIRL OF WORLD WAR II?	GRABLE
44	WHAT IS THE LAST NAME OF THE PRESIDENT OF THE UNITED STATES OF AMERICA DURING THE GULF WAR?	BUSH
45	WHAT IS THE LAST NAME OF THE SCIENTIST WHO DISCOVERED PENICILLIN?	FLEMMING
46	WHAT IS THE LAST NAME OF THE SECOND UNITED STATES PRESIDENT?	ADAMS
47	WHAT IS THE LAST NAME OF THE UNION GENERAL WHO DEFEATED THE CONFEDERATE ARMY AT THE CIVIL WAR BATTLE OF GETTYSBURG?	MEADE
48	WHAT IS THE LAST NAME OF THE VILLAINOUS CAPTAIN IN THE STORY ‘PETER PAN’?	HOOK
49	WHAT IS THE LONGEST RIVER IN EUROPE?	VOLGA
50	WHAT IS THE LONGEST RIVER IN SOUTH AMERICA?	AMAZON
51	WHAT IS THE NAME FOR THE ASTRONOMICAL BODIES THAT ENTER THE EARTH’S ATMOSPHERE?	METEORS
52	WHAT IS THE NAME OF A DRIED GRAPE?	RAISIN
53	WHAT IS THE NAME OF A DRIED PLUM?	PRUNE
54	WHAT IS THE NAME OF A GIANT OCEAN WAVE CAUSED BY AN EARTHQUAKE?	TSUNAMI
55	WHAT IS THE NAME OF A NUMBER TWO WOOD IN GOLF?	BRASSIE
56	WHAT IS THE NAME OF AN AIRPLANE WITHOUT AN ENGINE?	GLIDER
57	WHAT IS THE NAME OF BATMAN’S BUTLER?	ALFRED
58	WHAT IS THE NAME OF DEER MEAT?	VENISON
59	WHAT IS THE NAME OF DOROTHY’S DOG IN ‘THE WIZARD OF OZ’?	TOTO
60	WHAT IS THE NAME OF SOCRATES’ MOST FAMOUS STUDENT?	PLATO

Table G.3: Questions (61-90) and answers for Experiment 11

No.	Question	Correct answer
61	WHAT IS THE NAME OF THE AUTOMOBILE INSTRUMENT THAT MEASURES MILEAGE?	ODOMETER
62	WHAT IS THE NAME OF THE BIRD THAT CANNOT FLY AND IS THE LARGEST BIRD ON EARTH?	OSTRICH
63	WHAT IS THE NAME OF THE BRIGHTEST STAR IN THE SKY EXCLUDING THE SUN?	SIRIUS
64	WHAT IS THE NAME OF THE CHINESE RELIGION FOUNDED BY LAO TSE?	TAOISM
65	WHAT IS THE NAME OF THE COMIC STRIP CHARACTER WHO EATS SPINACH TO INCREASE HIS STRENGTH?	POPEYE
66	WHAT IS THE NAME OF THE CRIME IN WHICH A BUILDING OR PROPERTY IS PURPOSELY SET ON FIRE?	ARSON
67	WHAT IS THE NAME OF THE DESERT PEOPLE WHO WANDER INSTEAD OF LIVING IN ONE PLACE?	NOMADS
68	WHAT IS THE NAME OF THE FOUNTAIN IN ROME INTO WHICH COINS ARE THROWN FOR GOOD LUCK?	TREVI
69	WHAT IS THE NAME OF THE FURRY ANIMAL THAT ATTACKS COBRA SNAKES?	MONGOOSE
70	WHAT IS THE NAME OF THE HORSE-LIKE ANIMAL WITH BLACK AND WHITE STRIPES?	ZEBRA
71	WHAT IS THE NAME OF THE INSTRUMENT USED TO MEASURE WIND SPEED?	ANEMOMETER
72	WHAT IS THE NAME OF THE ISLAND ON WHICH NAPOLEON WAS BORN?	CORSICA
73	WHAT IS THE NAME OF THE KIND OF CAT THAT SPOKE TO ALICE IN THE STORY 'ALICE'S ADVENTURES IN WONDERLAND'?	CHESIRE
74	WHAT IS THE NAME OF THE LARGEST DESERT ON EARTH?	ANTARTICA
75	WHAT IS THE NAME OF THE LARGEST OCEAN ON EARTH?	PACIFIC
76	WHAT IS THE NAME OF THE LIGHTEST WOOD KNOWN?	BALSA
77	WHAT IS THE NAME OF THE LIZARD THAT CHANGES ITS COLOR TO MATCH THE SURROUNDINGS?	CHAMELEON
78	WHAT IS THE NAME OF THE LONG SLEEP SOME ANIMALS GO THROUGH DURING THE ENTIRE WINTER?	HIBERNATION
79	WHAT IS THE NAME OF THE MOLTEN ROCK THAT RUNS DOWN THE SIDE OF A VOLCANO DURING AN ERUPTION?	LAVA
80	WHAT IS THE NAME OF THE MOUNTAIN RANGE IN WHICH MOUNT EVEREST IS LOCATED?	HIMALAYAS
81	WHAT IS THE NAME OF THE MOUNTAIN RANGE THAT SEPARATES ASIA FROM EUROPE?	URAL
82	WHAT IS THE NAME OF THE NAVIGATION INSTRUMENT USED AT SEA TO PLOT POSITION BY THE STARS?	SEXTANT
83	WHAT IS THE NAME OF THE OCEAN THAT IS LOCATED BETWEEN AFRICA AND AUSTRALIA?	INDIAN
84	WHAT IS THE NAME OF THE ORGAN THAT PRODUCES INSULIN?	PANCREAS
85	WHAT IS THE NAME OF THE PROCESS BY WHICH PLANTS MAKE THEIR FOOD?	PHOTOSYNTHESIS
86	WHAT IS THE NAME OF THE REMAINS OF PLANTS AND ANIMALS THAT ARE FOUND IN STONE?	FOSSILS
87	WHAT IS THE NAME OF THE RUBBER OBJECT THAT IS HIT BACK AND FORTH BY HOCKEY PLAYERS?	PUCK
88	WHAT IS THE NAME OF THE SEVERE HEADACHE THAT RETURNS PERIODICALLY AND OFTEN IS ACCOMPANIED BY NAUSEA?	MIGRAINE
89	WHAT IS THE NAME OF THE SHORT PLEATED SKIRT WORN BY MEN IN SCOTLAND?	KILT
90	WHAT IS THE NAME OF THE SMALL JAPANESE STOVE USED FOR OUTDOOR COOKING?	HIBACHI

Table G.4: Questions (91-120) and answers for Experiment 11

No.	Question	Correct answer
91	WHAT IS THE NAME OF THE SMALLEST BONE IN THE HUMAN BODY?	STAPES
92	WHAT IS THE NAME OF THE SPEAR LIKE OBJECT THAT IS THROWN DURING AN ATHLETICS COMPETITION?	JAVELIN
93	WHAT IS THE NAME OF THE SUBSTANCE DERIVED FROM A WHALE THAT IS USED TO MAKE PERFUME?	AMBERGRIS
94	WHAT IS THE NAME OF THE UNIT OF MEASUREMENT THAT REFERS TO A SIX-FOOT DEPTH OF WATER?	FATHOM
95	WHAT IS THE ONLY WORD THAT THE RAVEN SAYS IN EDGAR ALLEN POE'S POEM 'THE RAVEN'?	NEVERMORE
96	WHAT IS THE TERM FOR HITTING A VOLLEYBALL DOWN HARD INTO THE OPPONENT'S COURT?	SPIKE
97	WHAT IS THE UNIT OF ELECTRICAL POWER THAT REFERS TO A CURRENT OF ONE AMPERE AT ONE VOLT?	WATT
98	WHAT IS THE UNIT OF SOUND INTENSITY?	DECIBEL
99	WHAT ISLAND IS THE LARGEST IN THE WORLD EXCLUDING AUSTRALIA?	GREENLAND
100	WHAT KIND OF METAL IS ASSOCIATED WITH A 50TH WEDDING ANNIVERSARY?	GOLD
101	WHAT KIND OF POISON DID SOCRATES TAKE AS HIS EXECUTION?	HEMLOCK
102	WHAT WAS FRANK LLOYD WRIGHT'S PROFESSION?	ARCHITECT
103	WHAT WAS THE CAPITAL OF WEST GERMANY PRIOR TO REUNIFICATION IN 1990?	BONN
104	WHAT WAS THE LAST NAME OF BUFFALO BILL?	CODY
105	WHAT WAS THE LAST NAME OF THE FIRST EUROPEAN EXPLORER TO LAND IN NORTH AMERICA?	ERIKSON
106	WHAT WAS THE NAME OF TARZAN'S GIRLFRIEND?	JANE
107	WHAT WAS THE NAME OF THE APOLLO LUNAR MODULE THAT LANDED THE FIRST MAN ON THE MOON?	EAGLE
108	WHAT WAS THE NAME OF THE SHIP ABOARD WHICH CHARLES DARWIN TRAVELLED TO THE GALAPAGOS?	BEAGLE
109	WHAT WAS THE NAME OF THE SUPPOSEDLY UNSINKABLE SHIP THAT SUNK ON ITS MAIDEN VOYAGE IN 1912?	TITANIC
110	WHICH COUNTRY WAS THE FIRST TO USE GUNPOWDER?	CHINA
111	WHICH FLOWER IS A NATIONAL SYMBOL IN WALES?	DAFFODIL
112	WHICH PLANET WAS THE LAST TO BE DISCOVERED?	NEPTUNE
113	WHICH PRECIOUS GEM IS RED?	RUBY
114	WHICH SPORT USES THE TERMS 'GUTTER' AND 'ALLEY'?	BOWLING
115	WHICH SPORT USES THE TERMS 'STONES' AND 'BROOMS'?	CURLING
116	WHICH TYPE OF SNAKE DO ASIAN SNAKE-CHARMERS USE?	COBRA
117	WHO IS KNOWN AS 'THE FATHER OF GEOMETRY'?	EUCLID
118	WHO KILLED THE MINOTAUR ACCORDING TO GREEK MYTHOLOGY?	THESEUS
119	WHO WAS THE FIRST RULER OF THE HOLY ROMAN EMPIRE?	CHARLEMAGNE
120	WHO WAS THE GREEK GOD OF WAR?	ARES

Appendix H

H.1 General instructions

Presented at the start of the study.

“We are investigating how people learn and remember information. In this particular study we will be asking you general knowledge questions and providing you with correcting information if you make errors.

The study is divided into three stages taking place over three consecutive days. You must only take part in the study if you are prepared to participate on all three days. Please note that all of the stages must be completed at the same time of day. We’ll give you a couple of hours of leeway with this. So for example, if you started stage one at 11am, you should aim to start stage two at 11am the following day, however in order to give you some flexibility we will actually allow you to start the stage anytime between 9am and 1pm. The same would apply for stage three on the third day.

If any of this is unclear, please get in touch by clicking the ‘contact us’ button at the bottom of the screen. If you fail to complete a stage, or do not complete the stage during the allocated times, you will be unable to take part in the rest of the study. The time required for each stage is variable. We estimate that stage one will take approximately 30-45 minutes, stage two will take approximately 45-60 minutes and stage three will take approximately 30-45 minutes.

Each stage is slightly different but will generally require you to answer a series of general knowledge questions, for example: ‘Who invented the light bulb?’ If you provide an incorrect answer to one of these questions, then we will teach you the correct answer. All information provided during the study will be factual. Further instructions will be provided throughout the study and if you encounter any problems you can contact us using the button at the bottom of the screen.

Please note that this is a scientific study. You must read all instructions carefully, take your time, and give your undivided attention to the task at all times. In return, you will be contributing valuable data to a scientific research study, and receive monetary compensation for your time. It is absolutely vital that you do not break any of the study rules. For example, during the study you must not look up the answers to any of the questions that we ask. The payment you receive is not related to the number of questions that you answer correctly, but is contingent on you strictly following the instructions we provide. Upon completion of the study, the research team will check your responses and then send you information about how to obtain your payment."

H.2 Study rules

Presented at the start of each stage.

"Please read and adhere to the following rules. In order to maintain data quality participants who do not follow these rules may face early exclusion from the study:

- Find a quiet place to do the study where you are unlikely to be distracted and where you have a stable Internet connection. Turn off your mobile phone, e-mail, and any other devices/software that might distract you.
- You should set aside enough time to complete the stage in a single sitting. You must not leave the study website during a stage. Do not use the 'back', 'forward', or 'reload' buttons on your browser or visit other websites during the study as doing so will cause your data to be lost and you will be unable to continue the study.
- If you leave the study webpage you will receive a warning message. Warnings can also be triggered by software on your computer that sends you notifications (e.g., about an e-mail) so it is advisable to turn off such applications for the duration of the study. If you receive too many warnings you will be excluded from the remainder of the study.
- is absolutely vital that you do not attempt to look up the answers to the questions until after the completion of the study. This is very important! We ask this to ensure

that all participants receive a similar experience during the study. You should also avoid writing down information about the questions or answers.

- Occasionally you will encounter a question that asks you to type in a short, 5-digit code. Be careful to enter this code correctly, we are checking that you are paying attention!"

H.3 Stage 1 instructions

Presented at the start of stage 1.

"Welcome to stage 1. During this stage we will test your knowledge about general kinds of information, for example, 'Who invented the light bulb?' You will be asked a question, and you will type your best guess in the space provided below the question. The answer will always consist of exactly one word and will never be longer than one word.

The questions vary greatly in difficulty such that you will probably be able to answer some easily, while others will be harder, and still others you may not know at all. If, after searching your memory, you are sure you don't know the answer, then you should click the 'SKIP' button.

If you think you know the answer, enter your response into the box provided. You should also provide a rating of your confidence in that response using the slider, which ranges from 1% to 100%.

Don't worry if you don't know the exact spelling of a word, our software can usually work out what you mean if you get close enough - just aim to be as accurate as you can. When you have entered a response and provided a confidence rating, click the 'SUBMIT' button to move on to the next question. When this happens you will notice that the progress bar at the bottom of the screen changes to show how much of this stage you have completed."

H.4 Stage 2 instructions

Presented at the start of stage 2.

“Welcome to stage 2. During this stage we will return to some of the questions that you answered on the stage one quiz and review the answers that you gave. Some of the questions will be ones that you answered correctly on the stage one quiz and others will be questions that you answered incorrectly.

If you gave the correct answer during stage 1, we will inform you of this. If you gave an incorrect response on stage 1 then we will provide you with the correct answer which you should try to remember.

Sometimes you will be asked to recall or restudy the responses you gave during stage 2 before you find out whether that response was correct or incorrect. On recall trials, you will be asked to enter your stage 1 response. On restudy trials your stage 1 response will be displayed on the screen. Recall and restudy trials will occur regardless of whether your stage 1 response was correct or incorrect.

You will also encounter a series of rest trials where you will be given a few moments to relax before we continue with the study. You should not leave the screen during this time and be ready to start the next trial.

Each trial will last for 20 seconds and you will automatically proceed to the next trial. A timer will be displayed in the bottom right of the screen to show you how long you have left on each trial.”

H.5 Stage 3 instructions

Presented at the start of stage 3.

“Welcome to stage 3. During this stage we will return to some of the questions that you have encountered during the study. The stage will be divided into three parts and we will begin with the instructions for part one.

As each question is presented search your memory hard and type in the correct answer. You should also provide a rating of your confidence in that response using the slider, which ranges from 1% to 100%. When you are happy with your response, click the ‘SUBMIT’ button.

Part two will automatically follow on from part one. We will cycle through each of the questions one final time and you should follow the on-screen instructions which will tell you which responses you need to provide. After part two you will be asked a series of general questions about the study and then you will be finished!"

Appendix I

Instruction and attention checks for Experiment 12

Two instruction/attention checks (see Oppenheimer et al., 2009) were employed to identify and exclude any participants who were not reading the instructions carefully or paying attention. In the first check during the instructions stage, participants were asked how often they use the Internet and given a series of options (Daily, Weekly, Monthly, Infrequently, Never). Within the main instruction block, they were told to ignore this question and answer 'Never', a response that is technically impossible (as the participant was using the Internet to take part in the study). In the second check during the question stage, participants were asked "Approximately how many Twitter messages do you think you have just read?" (there were actually 18) and told to respond with "more than 40" amongst other more realistic options (0-10, 11-20, 21-30, 31-40).

Appendix J

Instructions for Experiment 12

J.1 General instructions

At beginning of study.

“We are interested in how people understand and remember news reports. In this experiment you will see a news report presented in a series of short messages in the form of ‘Tweets’ from the social media platform Twitter. The Tweets will only be presented once and you will not be able to go back and re-read previous Tweets. Therefore you should take as long as you need to read each Tweet. Each Tweet will appear for a *minimum* of 5 seconds and then remain on screen until you click the >> button. After reading all of Tweets from this news report you will complete a short arithmetic test. Finally, you will be asked a series of questions about the event that you have been reading about.”

J.2 Inference question instructions

Included with each question.

“Please respond in as much detail as possible and provide reasons for your answer.”

Appendix K

Narratives used in Experiment 12

Message types are either filler, (no) false information, (no) retraction, or auxiliary detail. Note that in the experiment reported here, auxiliary details were always consistent with the alternative theme. The false theme details are for use in subsequent experiments.

K.1 Warehouse fire narrative

Message 1 [filler]: Emergency services have received a call reporting a fire at the premises of a large warehouse located near the city's main football stadium

Message 2 [filler]: The alarm was first raised by a security guard who saw smoke and called the emergency services

Message 3 [filler]: The fire is already spreading rapidly. An off-duty police officer walking nearby was the first to reach the scene and has begun an evacuation

Message 4 [filler]: Based on a recent company audit, the warehouse stock is estimated to be worth around one hundred thousand pounds

Message 5 [filler]: Two warehouse workers have been taken to hospital after sustaining injuries during the fire

Message 6 [false information]: Statement from Police Detective Lucas: "A storeroom contained cans of oil paint and pressurised gas cylinders near faulty electrical wiring"

OR: [no false information]: Statement from Police Detective Lucas: "We are working with the fire services to rapidly get the situation under control"

Message 7 [filler]: A small crowd of onlookers has gathered on a nearby pedestrian bridge to watch as the fire consumes the warehouse

Message 8 [filler]: Emergency services say they are grateful that the alarm was raised immediately after smoke was seen

Message 9 [auxiliary detail1 – false theme consistent]: There is a strong smell of burning and firefighters are reporting thick, oily smoke, sheets of flames and an intense heat

OR: [auxiliary detail1 – alternative theme consistent]: There is a strong smell of burning but firefighters are reporting that there is relatively little heat and most of the smoke is subsiding

Message 10 [filler]: Emergency services are reporting that the evacuation is now complete

Message 11 [auxiliary detail2 – false theme consistent]: A number of loud explosions in the storeroom have endangered the firefighters

OR: [auxiliary detail2 – alternative theme consistent]: Despite two small explosions in the staff kitchen, firefighters now have the fire under control

Message 12 [filler]: The warehouse owner says he is worried the fire will destroy his stock

Message 13 [auxiliary detail3 – false theme consistent]: A firefighter has emerged from the storeroom with what looks to be frayed electrical cables

OR: [auxiliary detail3 – alternative theme consistent]: A firefighter has emerged from the staff kitchens with what looks to be a burnt out saucepan

Message 14 [filler]: The warehouse workers who were taken to hospital have only sustained minor injuries and should make a full recovery

Message 15 [auxiliary detail4 – false theme consistent]: A warehouse worker was recently disciplined for breaching storeroom health and safety regulations regarding flammable material storage

OR: [auxiliary detail4 – alternative theme consistent]: A warehouse worker was recently disciplined for breaching health and safety regulations in the staff kitchens

Message 16 [filler]: Warehouse workers who did not sustain injuries are gathered in a nearby café waiting for news about the extent of the damage.

Message 17 [retraction]: Police Detective Lucas has released a new statement: “the storeroom did not contain paint cans and gas cylinders, it was actually empty”

OR: [no retraction]: Police Detective Lucas has released a new statement: “the fire is now under control and the area will soon be declared safe”

Message 18 [filler]: The fire service is reporting that the fire has been extinguished, but all of the warehouse stock has been destroyed

K.2 Zoo breakout narrative

Message 1 [filler]: We have received a report from the press office of a local zoo that a group of Roloway monkeys have escaped from their cage

Message 2 [filler]: The alarm was first raised by a staff member who found the cage empty when he went to deliver the lunchtime feed

Message 3 [filler]: Security teams are taking steps to lock down the premises after being notified of the breakout via the public tannoy system.

Message 4 [filler]: The Roloway monkeys are an endangered species imported from West Africa and are one of the zoo’s most valuable species

Message 5 [filler]: A family visiting the zoo say they saw three monkeys rummaging through the bins behind the zoo café and devouring a crate of lemons

Message 6 [false information]: Statement from Zoo Manager Kat Barns: “A work experience student left the monkey cage door open this morning after the morning feed”

OR: [no false information]: Statement from Zoo Manager Kat Barns: “We are working rapidly to locate and secure the monkeys.”

Message 7 [filler]: The zoo is one of the city’s most regularly visited tourist attractions and recently received an award for its outstanding insect collection

Message 8 [filler]: Zoo Head of Security has praised the individual who first raised the alarm this morning

Message 9 [auxiliary detail1 – false theme consistent]: The police recently visited the zoo to discuss incidents of theft and anti-social behaviour amongst the work experience students

OR: [auxiliary detail1 – alternative theme consistent]: The police recently visited the zoo after work experience students reported seeing a suspicious individual near the zoo's perimeter fence

Message 10 [filler]: Zoo Head of Security has stated that his staff rapidly locked down the area after hearing of the breakout

Message 11 [auxiliary detail2 – false theme consistent]: Staff members have recently raised concerns with management about work experience students being absent from compulsory training sessions

OR: [auxiliary detail2 – alternative theme consistent]: Staff members have recently raised concerns with management about the close proximity of the monkey enclosure to a perimeter fence

Message 12 [filler]: Several Roloway monkey experts have arrived on the scene, concerned for the safety of these valuable animals

Message 13 [auxiliary detail3 – false theme consistent]: The Zoo's Head of Security says an investigation into the zoo's work experience scheme has been launched

OR: [auxiliary detail3 – alternative theme consistent]: The Zoo's Head of Security says an investigation has been launched after a hole was found in the zoo's perimeter fence near the monkey cage

Message 14 [filler]: The family who saw the monkeys behind the zoo café told our reporter that the monkeys ate an entire crate of fruit and then quickly moved on

Message 15 [auxiliary detail4 – false theme consistent]: Police are examining CCTV footage that shows a group of work experience students arriving for the morning shift

OR: [auxiliary detail4 – alternative theme consistent]: Police are examining CCTV footage that shows a man wearing an "animal rights" shirt near the perimeter fence this morning

Message 16 [filler]: The police are contacting local residents by telephone, asking them not to approach the monkeys and urging them to report any sightings

Message 17 [retraction]: Statement from Zoo Manager Kat Barns: "the monkey cage was not left open by a work experience student, it was locked by a staff member"

OR: [no retraction]: Zoo Manager Kat Barns has stated that most of the monkeys have been recovered and opening hours will commence as normal tomorrow.

Message 19 [filler]: The zoo is reporting that all of the monkeys have now been located and secured in a temporary enclosure. A full investigation is expected.

Appendix L

Recognition memory tests used in Experiment 12

Instruction:

“Which of these Tweets have you seen before? You must choose one of the Tweets, even if you cannot remember seeing either of them.”

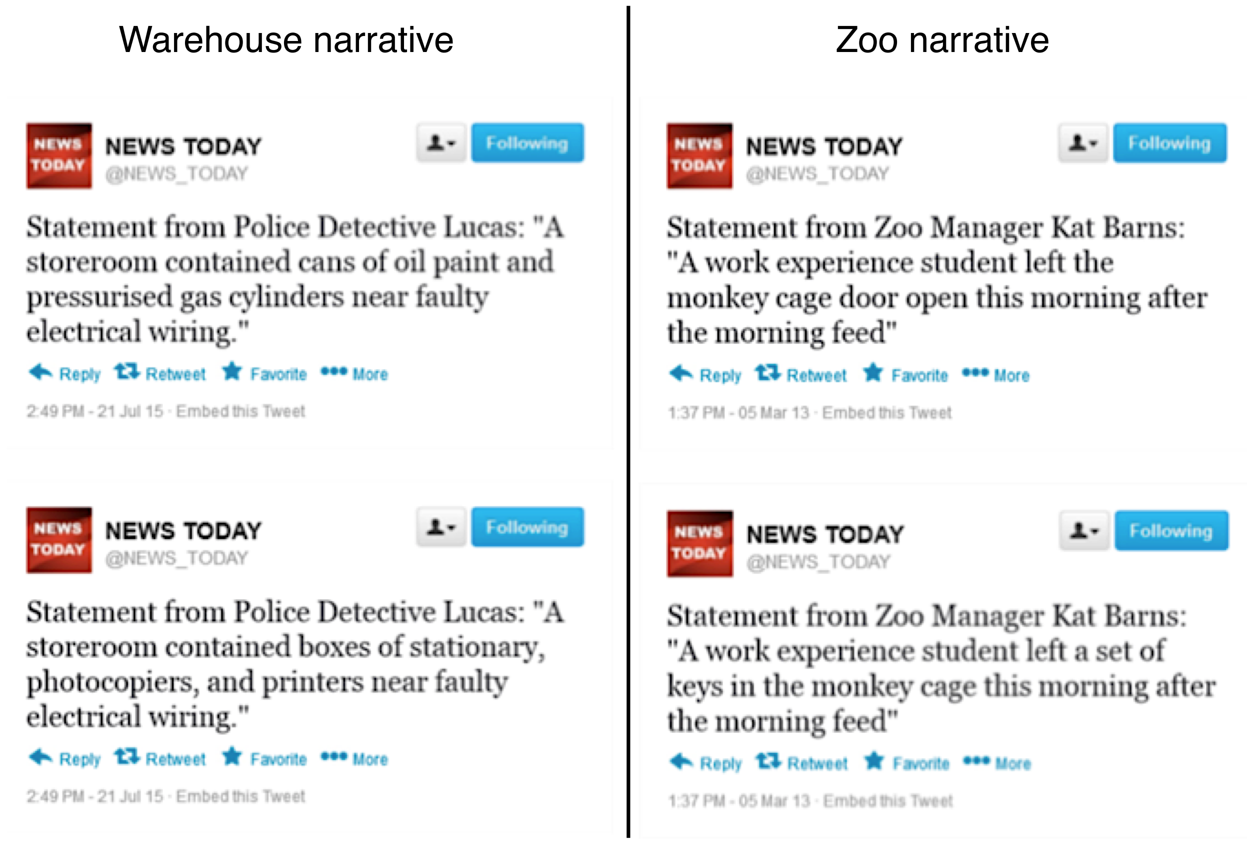


Figure L.1: Recognition memory tests used in Experiment 12. Recognition memory test options: original false information message (top) and lure (bottom). Position counter-balanced across participants.

Appendix M

Coding framework used in Experiment 12

M.1 Coding criteria for inference questions

Two dependent measures are being extracted from participants' responses to the inference questions:

- 'References to false information'
- 'References to alternative theme'

For each question, the coder will evaluate the participant's response and make a judgment as to whether it refers to the false information theme, the alternative theme, or neither.

- If the participant refers to neither theme, they will score 0 on both dependent measures.
- If the participant makes a causal reference consistent with the false information theme they will score 1 on the 'references to false information' measure.
- If the participant makes a causal reference consistent with the alternative information theme they will score 1 on the 'references to alternative theme' measure.
- If the participant makes causal references to both themes will score 0.5 on each measure.

The maximum score that can be achieved for each measure on each question is 1, even if multiple references are made in the same response. Thus, the maximum score for each measure for all inference questions is 8.

A causal reference is one in which the participant explicitly states, or strongly implies, that the warehouse fire was caused by gas and oil paints/unattended hot stove or the zoo breakout was caused by the cage being left open by a work experience student/an activist cutting a hole in the enclosure.

Table M.1: Zoo story inference questions and example responses.

Question	Example false information theme answer (to gain a score of 1)	Example alternative theme answer (to gain a score of 1)
Precisely how might the CCTV footage relate to the monkey's escape?	The footage showed that the students were at the zoo during the time of the escape	The footage showed that an animal rights activist was at the zoo during the time of the escape
What do you think will be the outcome of the Head of Security's investigation?	One of the students left the cage open when they were giving the monkeys their morning feed	An animal rights activist cut a hole in the fence of the monkey enclosure
How could the monkey escape have been avoided?	If the individuals on work experience were being kept under observation,while carrying out feeding then the enclosure wouldn't have been left open	If the fence was more secure then the animal rights activist may have been stopped
Are there any reasons why the police might wish to interview the work experience students again?	Yes – to ask if they left the cage open deliberately	Yes – to ask for a description of the animal rights activist they saw near the fence
Do you think zoo management should have listened more closely to the concerns of their staff?	Yes as a work experience student was to blame and this could have been avoided if earlier complaints were listened to	Yes as they had noticed that the monkey enclosure was too close to the perimeter fence
Who might the police consider arresting for their involvement in the monkey's escape?	A student as they might have done it on purpose	The animal rights activist as they broke into the zoo
What responsibility does the zoo's Head of Security hold for the monkey escape?	He should have been more careful when deciding who gets employed for work experience	The security arrangements were not good enough to stop the animal rights activist
How did the monkeys escape from the zoo?	A work experience student left the cage door open	An animal rights activist cut a hole in the monkey enclosure

Table M.2: Warehouse story inference questions and example responses.

Question	Example false information theme answer (to gain a score of 1)	Example alternative theme answer (to gain a score of 1)
Where is the fire most likely to have originated from?	The storeroom containing the paint and gas cylinders	The saucepan in the staff kitchens
What were the possible causes of the workers' injuries?	Inhaling toxic fumes from the oil paints or hit by exploding gas cylinders	Touching the hot pan in the kitchen
What evidence is there of carelessness?	Storing the pressurised cans and paints near the faulty wiring	Leaving the pan unattended on the stove
How could this fire have been avoided?	Careful storage of flammable materials	Not leaving saucepans unattended
In what way should the health and safety regulations have prevented the fire?	They gave advice about the appropriate storage of flammable materials	They gave advice about cooking safely in the staff kitchens
Do you think any workers should be disciplined for their role in the fire?	Yes for storing the toxic and flammable chemicals carelessly	Yes for leaving the saucepan unattended on the stove
What was the possible cause of the strong burning smell?	From the oil paints and gas cylinders	From the burnt saucepan
What was the most likely cause of the fire?	Flammable materials – the oil paints and gas cylinders – set on fire by the faulty electrical wiring	The saucepan being left unattended and catching fire

M.2 Coding criteria for filler questions

One dependent measure is being extracted from participants' responses to the filler questions:

- 'Recall of fillers'

For each question, the coder will evaluate the participant's response and make a judgment as to whether it approximates the correct answer.

- If the participant's response is similar to the correct answer they will score 1.
- If their response is incorrect they will score 0.

The maximum score that can be achieved on each question is 1. Thus, the maximum score for all filler questions is 8.

Table M.3: Zoo story filler questions and correct responses.

Question	Correct response
Where will the monkeys be kept until the investigation into the escape is complete?	A temporary enclosure
Who sent reports about the missing monkeys to the press?	The zoo's press office
Why did the zoo recently receive an award?	For its insect collection
How did the police contact local residents to inform them about the monkeys' escape?	By telephone
Who was the first to notice the monkeys were missing?	A staff member
By what method were zoo staff alerted about the missing monkeys?	Over the main tannoy system
Which continent did the monkey's inhabit in the wild?	Africa
Which fruit were the monkeys seen eating at the rear of the zoo café?	Lemons

Table M.4: Warehouse story filler questions and correct responses.

Question	Correct response
Who first raised the alarm?	A security guard
Approximately how much was the warehouse stock worth?	Around one hundred thousand pounds
How many warehouse workers were taken to hospital?	Two
How did the first police officer arrive on the scene?	He was walking nearby
Where did onlookers gather to watch the fire?	A bridge
How much of the warehouse stock was destroyed?	All of the stock was destroyed
What sporting venue was the warehouse located near?	Football stadium
Where did uninjured warehouse workers gather while they waited for news about the fire?	Cafe

M.3 Coding criteria for retraction awareness questions

One dependent measure is being extracted from participants' responses to the retraction awareness questions: 'Retraction awareness'.

The two questions are:

1. What was the point of the second message from the zoo's Assistant Manager Kat Barns? [zoo story]

OR What was the point of the second message from Police Detective Lucas? [warehouse story]

2. Were you aware of any corrections or contradictions in the messages that you read? Answers suggesting awareness of a retraction will receive a score of 1. There is a maximum score of 1 for each question, thus the maximum retraction awareness score is 2.

REFERENCES

Adams, D. (1979). *The Hitchhiker's Guide to the Galaxy*. London: Pan Books.

Agren, T. (2014). Human reconsolidation: A reactivation and update. *Brain Research Bulletin*, *105*, 70–82.

Agren, T., Engman, J., Frick, A., Björkstrand, J., Larsson, E.-M., Furmark, T., & Fredrikson, M. (2012). Disruption of reconsolidation erases a fear memory trace in the human amygdala. *Science*, *337*, 1550–1552.

Alberini, C. M. (2011). The role of reconsolidation and the dynamic process of long-term memory formation and storage. *Frontiers in Behavioral Neuroscience*, *5*, 1–10.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, *39*, 1037–1049.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*, 451–474.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186–197.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork (Ed.), *Memory* (pp. 237–313). London: Academic Press.

Anderson, R. C., & Watts, G. H. (1971). Response competition in the forgetting of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *10*, 29–34.

Anokhin, K. V., & Tiunova, A. A. (2002). Reminder effects: Reconsolidation or retrieval deficit? Pharmacological dissection with protein synthesis inhibitors following reminder for a passive-avoidance task in young chicks. *European Journal of Neuroscience*, *15*, 1759–1765.

Archer, E. J., & Underwood, B. J. (1951). Retroactive inhibition of verbal associations as a multiple function of temporal point of interpolation and degree of interpolated learning. *Journal of Experimental Psychology: General*, *42*, 283–290.

Auber, A., Tedesco, V., Jones, C. E., Monfils, M.-H., & Chiamulera, C. (2013). Post-retrieval extinction as reconsolidation interference: Methodological issues or boundary conditions? *Psychopharmacology*, *226*, 631–647.

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, *5*, 1–21.

Baker, K. D., McNally, G. P., & Richardson, R. (2013). Memory retrieval before or after extinction reduces recovery of fear in adolescent rats. *Learning & Memory*, *20*, 467–473.

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*, 666–678.

Bakker, M., Dijk, A. van, & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543–554.

Banker, G., Hunt, E., & Pagano, R. (1969). Evidence supporting the memory disruption hypothesis of electro-convulsive shock action. *Physiology & Behavior*, *4*, 895–899.

Barnes, J. M., & Underwood, B. J. (1959). Fate of first-list associations in transfer theory. *Journal of Experimental Psychology: General*, *58*, 97–105.

Barraco, R. A., & Stettner, L. J. (1976). Antibiotics and memory. *Psychological Bulletin*, *83*, 242–302.

Bekerian, D. A., & Bowers, J. M. (1983). Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 139–145.

Belli, R. F. (1989). Influences of misleading postevent information: Misinformation interference and acceptance. *Journal of Experimental Psychology: General*, *118*, 72–85.

Belli, R. F. (1993). Failure of interpolated tests in inducing memory impairment with final modified tests: Evidence unfavorable to the blocking hypothesis. *The American Journal of Psychology*, *106*, 407.

Belli, R. F., Windschitl, P. D., & McCarthy, T. T. (1992). Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 356–367.

Besnard, A., Caboche, J., & Laroche, S. (2012). Reconsolidation of memory: A decade of debate. *Progress in Neurobiology*, *99*, 61–80.

Beukelaar, T. T. de, Woolley, D. G., & Wenderoth, N. (2014). Gone for 60 seconds: Reactivation length determines motor memory degradation during reconsolidation. *Cortex*, *59*, 138–145.

Biedenkapp, J. C., & Rudy, J. W. (2004). Context memories and reactivation: Constraints on the reconsolidation hypothesis. *Behavioral Neuroscience*, *118*, 956–964.

Birnbaum, I. M. (1965). Long-term retention of first-list associations in the AB, AC paradigm. *Journal of Verbal Learning and Verbal Behavior*, *4*, 515–520.

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

Bjork, R. A. (1978). The updating of human memory. *The Psychology of Learning and Motivation*, *12*, 235–259.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness essays in honor of endel tulving* (pp. 1–12). Hillside, NJ.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *Learning processes to cognitive processes essays in honor of william k. estes* (pp. 35–67). Hillsdale, NJ.

Björkstrand, J., Agren, T., Frick, A., Engman, J., Larsson, E.-M., Furmark, T., & Fredrikson,

M. (2015). Disruption of memory reconsolidation erases a fear memory trace in the human amygdala: An 18-month follow-up. *PLoS ONE*, *10*, 1–8.

Blaiss, C. A., & Janak, P. H. (2007). Post-training, but not post-reactivation, administration of amphetamine and anisomycin modulates Pavlovian conditioned approach. *Neurobiology of Learning and Memory*, *87*, 644–658.

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

Boccia, M. M., Blake, M. G., Acosta, G. B., & Baratti, C. M. (2005). Memory consolidation and reconsolidation of an inhibitory avoidance task in mice: Effects of a new different learning task. *Neuroscience*, *135*, 19–29.

Bonto, M. A., & Payne, D. G. (1991). Role of environmental context in eyewitness memory. *The American Journal of Psychology*, *104*, 117–134.

Bos, M. G. N., Beckers, T., & Kindt, M. (2014). Noradrenergic blockade of memory reconsolidation: A failure to reduce conditioned fear responding. *Frontiers in Behavioral Neuroscience*, *8*, 412.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*, 80–99.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry*, *52*, 976–986.

Bouton, M. E. (2014). Why behavior change is difficult to sustain. *Preventive Medicine*, *68*, 29–36.

Bouton, M. E., & Bolles, R. C. (1979a). Contextual control of the extinction of conditioned fear. *Learning and Motivation*, *10*, 445–466.

Bouton, M. E., & Bolles, R. C. (1979b). Role of conditioned contextual stimuli in reinstatement of extinguished fear. *Journal of Experimental Psychology: Animal Behavior Processes*, *5*, 368–378.

Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear: tests for the associative value of the context. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 248–265.

Bouton, M. E., & Moody, E. W. (2004). Memory processes in classical conditioning. *Neuroscience and Biobehavioral Reviews*, *28*, 663–674.

Bouton, M. E., & Peck, C. A. (1992). Spontaneous recovery in cross-motivational transfer (counterconditioning). *Animal Learning & Behavior*, *20*, 313–321.

Bower, G. (1967). A multicomponent theory of the memory trace. *The Psychology of Learning and Motivation*, *1*, 229–324.

Bowman, L. L., & Zaragoza, M. S. (1989). Similarity of encoding context does not influence resistance to memory impairment following misinformation. *The American Journal of Psychology*, *102*, 249.

Bregman, N. J., Nicholas, T., & Lewis, D. J. (1976). Cue-dependent amnesia: Permanence and memory return. *Physiology & Behavior*, *17*, 267–270.

Briggs, G. E. (1954). Acquisition, extinction, and recovery functions in retroactive inhibition. *Journal of Experimental Psychology: General*, *47*, 285–293.

Brooks, D. C., Hale, B., Nelson, J. B., & Bouton, M. E. (1995). Reinstatement after counterconditioning. *Animal Learning & Behavior*, *23*, 383–390.

Brown, A. S. (1976). Spontaneous recovery in human learning. *Psychological Bulletin*, *83*, 321–338.

Brown, A. S. (2002). Consolidation theory and retrograde amnesia in humans. *Psychonomic Bulletin & Review*, *9*, 403–425.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539–576.

Bruer, J. (2007). Plasticity: On the level. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory concepts* (pp. 77–82). Oxford University Press.

Burnham, W. H. (1903). Retroactive amnesia: Illustrative cases and a tentative explanation. *The American Journal of Psychology*, *14*, 118–132.

Butler, A. C., Fazio, L. K., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin & Review*, *18*, 1238–1244.

Butler, A. C., Zaromb, F. M., Lyle, K. B., & Roediger, H. L. (2009). Using popular films to enhance classroom learning: The good, the bad, and the interesting. *Psychological Science*, *20*, 1161–1168.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 1–12.

Cahill, L., McGaugh, J. L., & Weinberger, N. M. (2001). The neurobiology of learning and memory: Some reminders to remember. *Trends in Neurosciences*, *24*, 578–581.

Cammarota, M., Bevilaqua, L. R. M., Medina, J. H., & Izquierdo, I. (2004). Retrieval does not induce reconsolidation of inhibitory avoidance memory. *Learning & Memory*, *11*, 572–578.

Capaldi, E. J., & Neath, I. (1995). Remembering and forgetting as context discrimination. *Learning & Memory*, *2*, 107–132.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642.

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General*, *116*, 38–49.

Censor, N., Dayan, E., & Cohen, L. G. (2014a). Cortico-subcortical neuronal circuitry associated with reconsolidation of human procedural memories. *Cortex*, *58*, 281–288.

Censor, N., Dimyan, M. A., & Cohen, L. G. (2010). Modification of existing human motor memories is enabled by primary cortical processing during memory reactivation. *Current Biology*, *20*, 1545–1549.

Censor, N., Horovitz, S. G., & Cohen, L. G. (2014b). Interference with existing memories alters offline intrinsic functional brain connectivity. *Neuron*, *81*, 69–76.

Cervantes, M. de. (1998). *Don Quixote*. New York: Random House.

Chamberlin, T. C. (1965). The method of multiple working hypotheses. *Science*, *148*, 754–759.

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*, 153–170.

Chan, J. C. K., & Langley, M. M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 248–255.

Chan, J. C. K., & LaPaglia, J. A. (2011). The dark side of testing memory: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: General*, *17*, 418–432.

Chan, J. C. K., & LaPaglia, J. A. (2013). Impairing existing declarative memory in humans by disrupting reconsolidation. *Proceedings of the National Academy of Sciences of The United States of America*, *110*, 9309–9313.

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science*, *20*, 66–73.

Chan, W. Y. M., Leung, H. T., Westbrook, R. F., & McNally, G. P. (2010). Effects of recent exposure to a conditioned stimulus on extinction of Pavlovian fear conditioning. *Learning & Memory*, *17*, 512–521.

Chandler, C. C. (1989). Specific retroactive interference in modified recognition tests: Evidence for an unknown cause of interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 256–265.

Chandler, C. C. (1991). How memory for an event is influenced by related events: Interference in modified recognition tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 115–125.

Chandler, C. C. (1993). Accessing related events increases retroactive interference in a matching recognition test. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 967–974.

Chandler, C. C., & Fisher, R. P. (1996). Retrieval processes and witness memory. In E. L. Bjork (Ed.), *Memory* (pp. 493–524). London.

Cherkin, A. (1972). Retrograde amnesia in the chick: Resistance to the reminder effect. *Physiology & Behavior*, *8*, 949–955.

Christiaansen, R. E., & Ochalek, K. (1983). Editing misleading information from memory: Evidence for the coexistence of original and postevent information. *Memory & Cognition*, *11*, 467–475.

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.

Coccoz, V., Sandoval, A. V., Stehberg, J., & Delorenzi, A. (2013). The temporal dynamics of enhancing a human declarative memory during reconsolidation. *Neuroscience*, *246*, 397–408.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cooper, R. M., & Koppenaal, R. J. (1964). Suppression and recovery of a one-trial avoidance response after a single ECS. *Psychonomic Science*, *1*, 303–304.

Costanzi, M., Cannas, S., Saraulli, D., Rossi-Arnaud, C., & Cestari, V. (2011). Extinction after retrieval: Effects on the associative and nonassociative components of remote contextual fear memory. *Learning & Memory*, *18*, 508–518.

Crowder, R. G. (1976). *Principles of learning and memory.* Michigan: Lawrence Erlbaum

Associates.

Crowder, R. G. (1989). Modularity and dissociations in memory systems. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness essays in honor of endel tulving* (pp. 274–294). Hillsdale, NJ: Erlbaum.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Davis, H. P., & Squire, L. R. (1984). Protein synthesis and memory: A review. *Psychological Bulletin*, *96*, 518–559.

Dawson, R. G., & McGaugh, J. L. (1969). Electroconvulsive shock effects on a reactivated memory trace: Further examination. *Science*, *166*, 525–527.

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

Debiec, J., LeDoux, J. E., & Nader, K. (2002). Cellular and systems reconsolidation in the hippocampus. *Neuron*, *36*, 527–538.

Dekker, S., Lee, N. C., Howard-Jones, P. A., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, *3*, 1–8.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of The United States of America*, *113*, 554–559.

DeVietti, T. L., & Holliday, J. H. (1972). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace: A replication. *Psychonomic Science*, *29*, 137–138.

Dewar, M., Cowan, N., & Della Sala, S. (2007). Forgetting due to retroactive interference in amnesia: A fusion of early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex*, *43*, 616–634.

Diekelmann, S., Büchel, C., Born, J., & Rasch, B. (2011). Labile or stable: Opposing consequences for memory when reactivated during waking and sleep. *Nature Neuroscience*, *14*, 381–386.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 1–17.

Dudai, Y. (1996). Consolidation: Fragility on the road to the engram. *Neuron*, *17*, 367–370.

Dudai, Y. (2000). The shaky trace. *Nature*, *406*, 686–687.

Dudai, Y. (2002). Molecular bases of long-term memories: A question of persistence. *Current Opinion in Neurobiology*, *12*, 211–216.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86.

Dudai, Y. (2009). Predicting not to predict too much: How the cellular machinery of memory anticipates the uncertain future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1255–1262.

Dudai, Y. (2012). The restless engram: Consolidations never end. *Annual Review of Neuroscience*, *35*, 227–247.

Dudai, Y., & Eisenberg, M. (2004). Rites of passage of the engram: Reconsolidation and the lingering consolidation hypothesis. *Neuron*, *44*, 93–100.

Duncan, C. P. (1949). The retroactive effect of electroshock on learning. *Journal of Comparative and Physiological Psychology*, *42*, 32–44.

Duvarci, S., & Nader, K. (2004). Characterization of fear memory reconsolidation. *The Journal of Neuroscience*, *24*, 9269–9275.

Ecker, U. K. H., Lewandowsky, S., & Apai, J. (2011a). Terrorists brought down the plane! No, actually it was a technical fault: Processing corrections of emotive information. *The*

Quarterly Journal of Experimental Psychology, *64*, 283–310.

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*, 1087–1100.

Ecker, U. K. H., Lewandowsky, S., Swire, B., & Chang, D. (2011b). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, *18*, 570–578.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634.

Eich, J. E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition*, *8*, 157–173.

Eich, J. E., Weingartner, H., Stillman, R. C., & Gillin, J. C. (1975). State-dependent accessibility of retrieval cues in the retention of a categorized list. *Journal of Verbal Learning and Verbal Behavior*, *14*, 408–417.

Einarsson, E. Ö., & Nader, K. (2012). Involvement of the anterior cingulate cortex in formation, consolidation, and reconsolidation of recent and remote contextual fear memory. *Learning & Memory*, *19*, 449–452.

Eisenberg, M., & Dudai, Y. (2004). Reconsolidation of fresh, remote, and extinguished fear memory in medaka: Old fears don't die. *European Journal of Neuroscience*, *20*, 3397–3403.

Eisenberg, M., Kobilo, T., Berman, D. E., & Dudai, Y. (2003). Stability of retrieved memory: Inverse correlation with trace dominance. *Science*, *301*, 1102–1104.

Exton-McGuinness, M. T. J., Lee, J. L., & Reichelt, A. C. (2015). Updating memories: The role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384.

Falls, W. A. (1998). Extinction: A review of theory and the evidence suggesting that memories are not erased with nonreinforcement. In W. O'Donohue (Ed.), *Learning and behavior* *therapy* (pp. 205–229). Boston: Learning; behavior therapy.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*, 891–904.

Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the jury disregard that information? The use of suspicion to reduce the prejudicial effects of pretrial publicity and inadmissible testimony. *Personality and Social Psychology Bulletin*, *23*, 1215–1226.

Fernandez, R. S., Bavassi, L., Forcato, C., & Pedreira, M. E. (2016). The dynamic nature of the reconsolidation process and its boundary conditions: Evidence based on human tests. *Neurobiology of Learning and Memory*, *130*, 202–212.

Feynman, R. (1974). Cargo cult science. In *Caltech commencement speech*.

Finn, B., & Roediger, H. L. (2011). Enhancing retention through reconsolidation: Negative emotional arousal following retrieval enhances later recall. *Psychological Science*, *22*, 781–786.

Finn, B., & Roediger, H. L. (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1665–1681.

Finn, B., Roediger, H. L., & Rosenzweig, E. (2012). Reconsolidation from negative emotional pictures: Is successful retrieval required? *Memory & Cognition*, *40*, 1031–1045.

Font, L., & Cunningham, C. L. (2012). Post-retrieval propranolol treatment does not modulate reconsolidation or extinction of ethanol-induced conditioned place preference. *Pharmacology, Biochemistry and Behavior*, *101*, 222–230.

Forcato, C., Argibay, P. F., Pedreira, M. E., & Maldonado, H. (2009). Human reconsolidation does not always occur when a memory is retrieved: The relevance of the reminder structure. *Neurobiology of Learning and Memory*, *91*, 50–57.

Forcato, C., Burgos, V. L., Argibay, P. F., Molina, V. A., Pedreira, M. E., & Maldonado, H. (2007). Reconsolidation of declarative memory in humans. *Learning & Memory*, *14*, 295–303.

Forcato, C., Fernandez, R. S., & Pedreira, M. E. (2013). The role and dynamic of strengthening in the reconsolidation process in a human declarative memory: What decides the fate of recent and older memories? *PLoS ONE*, *8*, 1–12.

Forcato, C., Rodríguez, M. L. C., & Pedreira, M. E. (2011). Repeated labilization-reconsolidation processes strengthen declarative memory in humans. *PLoS ONE*, *6*, 1–14.

Forcato, C., Rodríguez, M. L. C., Pedreira, M. E., & Maldonado, H. (2010). Reconsolidation in humans opens up declarative memory to the entrance of new information. *Neurobiology of Learning and Memory*, *93*, 77–84.

Forrester, W. E. (1970). Retroactive inhibition and spontaneous recovery in the AB, DC paradigm. *Journal of Verbal Learning and Verbal Behavior*, *9*, 525–528.

Fricchione, J., Greenberg, M. S., Spring, J. D., Wood, N. E., Mueller-Pfeiffer, C., Milad, M. R., ... Orr, S. P. (2016). Delayed extinction fails to reduce skin conductance reactivity to fear-conditioned stimuli. *The Psychology of Learning and Motivation*, *53*, 1343–1351.

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

Geller, A., & Jarvik, M. E. (1968a). Electroconvulsive shock induced amnesia and recovery. *Psychonomic Science*, *10*, 15–16.

Geller, A., & Jarvik, M. E. (1968b). The time relations of ECS induced amnesia. *Psychonomic Science*, *12*, 169–170.

Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis, a garden of forking paths, explains why many statistically significant comparisons don't hold up. *American Scientist*, *102*, 460–465.

Gelman, A., & Stern, H. (2006). The difference between significant and not significant is not itself statistically significant. *The American Statistician*, *60*, 328–331.

Gershman, S. J., Schapiro, A. C., Hupbach, A., & Norman, K. A. (2013). Neural context reinstatement predicts memory misattribution. *The Journal of Neuroscience*, *33*, 8590–8595.

Gerson, R., & Hendersen, R. W. (1978). Conditions that potentiate the effects of electroconvulsive shock administered 24 hours after avoidance training. *Animal Learning & Behavior*, *6*, 346–351.

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, *7*, 562–571.

Gisquet-Verrier, P., Lynch, J. F., Cutolo, P., Toledano, D., Ulmen, A., Jasnow, A. M., & Riccio, D. C. (2015). Integration of new information with active memory accounts for retrograde amnesia: A challenge to the consolidation/reconsolidation hypothesis? *The Journal of Neuroscience*, *35*, 11623–11633.

Glickman, S. E. (1961). Perseverative neural processes and consolidation of the memory trace. *Psychological Bulletin*, *58*, 218–233.

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

Gold, P. E. (2006). The many faces of amnesia. *Learning & Memory*, *13*, 506–514.

Gold, P. E., & King, R. A. (1972). Amnesia: Tests of the effect of delayed footshock-electroconvulsive shock pairings. *Physiology & Behavior*, *8*, 797–800.

Gold, P. E., & King, R. A. (1974). Retrograde amnesia: Storage failure versus retrieval failure. *Psychological Review*, *81*, 465–469.

Golkar, A., Bellander, M., Olsson, A., & Öhman, A. (2012). Are fear memories erasable? Reconsolidation of learned fear with fear-relevant and fear-irrelevant stimuli. *Frontiers in Behavioral Neuroscience*, *6*, 1–10.

Gordon, L. T., & Thomas, A. K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory & Cognition*, *42*, 186–197.

Gordon, L. T., Thomas, A. K., & Bulevich, J. B. (2015). Looking for answers in all the wrong places: How testing facilitates learning of misinformation. *Journal of Memory and Language*, *83*, 140–151.

Gordon, W. C. (1983). The malleability of memory in animals. In *Animal cognition and behavior* (pp. 399–426). Amsterdam: Elsevier Science.

Gordon, W. C., & Mowrer, R. R. (1980). An extinction trial as a reminder treatment following electroconvulsive shock. *Animal Learning & Behavior*, *8*, 363–367.

Gordon, W. C., & Spear, N. E. (1973). The effects of strychnine on recently acquired and reactivated passive avoidance memories. *Physiology & Behavior*, *10*, 1071–1075.

Greitemeyer, T. (2013). Article retracted, but the message lives on. *Psychonomic Bulletin & Review*, *21*, 557–561.

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, *2*, 201–209.

Gust, D. A., Darling, N., Kennedy, A., & Schwartz, B. (2008). Parents with doubts about vaccines: Which vaccines and reasons why. *Pediatrics*, *122*, 718–725.

Haaker, J., Golkar, A., Hermans, D., & Lonsdorf, T. B. (2014). A review on human reinstatement studies: An overview and methodological challenges. *Learning & Memory*, *21*, 424–440.

Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology*, *61*, 141–167.

Hardwicke, T. E., & Shanks, D. R. (2016). Reply to Walker and Stickgold: Proposed boundary conditions on memory reconsolidation will require empirical verification. *Proceedings of the National Academy of Sciences of The United States of America*, *113*, 1–2.

Hardwicke, T. E., Jameel, L., Jones, M., Walczak, E. J., & Magis-Weinberg, L. (2014). Only human: Scientists, systems, and suspect statistics. *Opticon1826*, *16*, 1–12.

Hardwicke, T. E., Taqi, M., & Shanks, D. R. (2016). Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proceedings of the National Academy of Sciences of The United States of America*, *113*, 5206–5211.

Harris, J. A., Jones, M. L., Bailey, G. K., & Westbrook, R. F. (2000). Contextual control over conditioned responding in an extinction paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 174–185.

Hart, P. S., & Nisbet, E. C. (2012). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, *39*, 701–723.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. New York: John Wiley & Sons.

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*, 193–233.

Herz, M. J., & Peeke, H. (1968). ECS-produced retrograde amnesia: Permanence vs. recovery over repeated testing. *Physiology & Behavior*, *3*, 517–521.

Hinderliter, C. F., Webster, T., & Riccio, D. C. (1975). Amnesia induced by hypothermia as a function of treatment-test interval and recooling in rats. *Animal Learning & Behavior*, *3*, 257–263.

Hintzman, D. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Howard, G. S., Maxwell, S. E., & Fleming, K. J. (2000). The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods*, *5*, 315–332.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context.

Journal of Mathematical Psychology, *46*, 269–299.

Howard-Jones, P. A. (2014). Neuroscience and education: Myths and messages. *Nature Reviews Neuroscience*, *15*, 817–824.

Hui, K., & Fisher, C. E. (2014). The ethics of molecular memory modification. *Journal of Medical Ethics*, *0*, 1–6.

Hupbach, A. (2015). Retrieval practice does not safeguard memories from interference-based forgetting. *Learning and Motivation*, *49*, 23–30.

Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: Updating or source confusion? *Memory*, *17*, 502–510.

Hupbach, A., Gomez, R., & Nadel, L. (2011). Episodic memory updating: The role of context familiarity. *Psychonomic Bulletin & Review*, *18*, 787–797.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, *14*, 47–53.

Hupbach, A., Hardt, O., Gomez, R., & Nadel, L. (2008). The dynamics of memory: Context-dependent updating. *Learning & Memory*, *15*, 574–579.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e124.

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, *7*, 645–654.

Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Medicine*, *11*, e1001747–6.

Ishii, D., Matsuzawa, D., Matsuda, S., Tomizawa, H., Sutoh, C., & Shimizu, E. (2015). An isolated retrieval trial before extinction session does not prevent the return of fear. *Behavioural Brain Research*, *287*, 139–145.

James, E. L., Bonsall, M. B., Hoppitt, L., Tunbridge, E. M., Geddes, J. R., Milton, A. L., &

Holmes, E. A. (2015). Computer game play reduces intrusive memories of experimental trauma via reconsolidation-update mechanisms. *Psychological Science*, *26*, 1201–1215.

Jamieson, J. L., & Albert, D. J. (1970). Amnesia from ECS: The effect of pairing ECS and footshock. *Psychonomic Science*, *18*, 14–15.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532.

Johnson, D. C., & Casey, B. J. (2015). Extinction during memory reconsolidation blocks recovery of fear in adolescents. *Scientific Reports*, *5*, 8863–5.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1420–1436.

Johnson, H. M., & Seifert, C. M. (1999). Modifying mental representations: Comprehending corrections. In S. R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representations during reading* (pp. 303–318). Mahwah, NJ.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

Johnson-Laird, P. N. (2012). Mental models and consistency. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency a unifying concept in social psychology* (pp. 1–30). New York: Guildford Press.

Jones, B. J., Bukoski, E., Nadel, L., & Fellous, J.-M. (2012). Remaking memories: Reconsolidation updates positively motivated spatial memory in rats. *Learning & Memory*, *19*, 91–98.

Jones, B. J., Pest, S. M., Vargas, I. M., Glisky, E. L., & Fellous, J.-M. (2015). Contextual reminders fail to trigger memory reconsolidation in aged rats and aged humans. *Neurobiology of Learning and Memory*, *120*, 7–15.

Jones, C. E., Ringuet, S., & Monfils, M.-H. (2013). Learned together, extinguished apart:

Reducing fear to complex stimuli. *Learning & Memory*, *20*, 674–685.

Josselyn, S. A., Köhler, S., & Frankland, P. W. (2015). Finding the engram. *Nature Reviews Neuroscience*, *16*, 521–534.

Judge, M. E., & Quartermain, D. (1982). Characteristics of retrograde amnesia following reactivation of memory in mice. *Physiology & Behavior*, *28*, 585–590.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217.

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., … Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, *14*, 1–15.

Kindt, M., & Soeter, M. (2013). Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. *Biological Psychology*, *92*, 43–50.

Kindt, M., Soeter, M., & Vervliet, B. (2009). Beyond extinction: Erasing human fear responses and preventing the return of fear. *Nature Neuroscience*, *12*, 256–258.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr., R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. A. (2014). Investigating variation in replicability: A Many Labs replication project. *Social Psychology*, *45*, 142–152.

Klucken, T., Kruse, O., Schweckendiek, J., Kuepper, Y., Mueller, E. M., Hennig, J., & Stark, R. (2016). No evidence for blocking the return of fear by disrupting reconsolidation prior to extinction learning. *Cortex*, *79*, 112–122.

Koppenaal, R. J. (1963). Time changes in the strengths of AB, A-C lists; Spontaneous recovery? *Journal of Verbal Learning and Verbal Behavior*, *2*, 310–319.

Koppenaal, R. J., Jagoda, E., & Cruce, J. (1967). Recovery from ECS-produced amnesia following a reminder. *Psychonomic Science*, *9*, 293–294.

Kraemer, P. J., & Golding, J. M. (1997). Adaptive forgetting in animals. *Psychonomic Bulletin & Review*, *4*, 480–491.

Kredlow, M. A., & Otto, M. W. (2015). Interference with the reconsolidation of trauma-related memories in adults. *Depression and Anxiety*, *32*, 32–37.

Kredlow, M. A., Unger, L. D., & Otto, M. W. (2015). Harnessing reconsolidation to weaken fear and appetitive memories: A meta-analysis of post-retrieval extinction effects. *Psychological Bulletin*, *142*, 314–336.

Kritchevsky, M., Zouzounis, J., & Squire, L. R. (1997). Transient global amnesia and functional retrograde amnesia: Contrasting examples of episodic memory loss. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*, 1747–1754.

Kroes, M. C. W., Tendolkar, I., Wingen, G. A. van, Waarde, J. A. van, Strange, B. A., & Fernández, G. (2014). An electroconvulsive therapy procedure impairs reconsolidation of episodic memories in humans. *Nature Neuroscience*, *17*, 204–206.

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology*, *68*, 565–579.

Lakatos, I. (1978). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge: Cambridge University Press.

Lakens, D., & Evers, E. R. K. (2014). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, *9*, 278–292.

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. *The Behavioral and Brain Sciences*, *38*, e1.

Lattal, K. M., & Abel, T. (2004). Behavioral impairments caused by injections of the protein synthesis inhibitor anisomycin after contextual retrieval reverse with time. *Proceedings of the National Academy of Sciences of The United States of America*, *101*, 4667–4672.

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 years of consolidation: Remembering

Müller and Pilzecker. *Learning & Memory*, *6*, 77–87.

Lee, J. L. (2009). Reconsolidation: Maintaining memory relevance. *Trends in Neurosciences*, *32*, 413–420.

Lewandowsky, S., Ecker, U. K. H., Farrell, S., & Brown, G. D. A. (2012a). Models of cognition and constraints from neuroscience: A case study involving consolidation. *Australian Journal of Psychology*, *64*, 37–45.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012b). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*, 106–131.

Lewis, D. J. (1976). A cognitive approach to experimental amnesia. *The American Journal of Psychology*, *89*, 51–80.

Lewis, D. J. (1979). Psychobiology of active and inactive memory. *Psychological Bulletin*, *86*, 1054–1083.

Lewis, D. J., & Maher, B. A. (1965). Neural consolidation and electroconvulsive shock. *Psychological Review*, *72*, 225–239.

Lewis, D. J., & Maher, B. A. (1966). Electroconvulsive shock and inhibition: Some problems considered. *Psychological Review*, *73*, 388–392.

Lewis, D. J., Bregman, N. J., & Mahan, J. J. (1972). Cue-dependent amnesia in rats. *Journal of Comparative and Physiological Psychology*, *81*, 243–247.

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1077–1083.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*, 349–358.

Loftus, E. F. (1979a). *Eyewitness Testimony.* Cambridge, MA: Havard University Press.

Loftus, E. F. (1979b). The malleability of human memory: Information introduced after we view an incident can transform memory. *American Scientist*, *67*, 312–320.

Loftus, E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361–366.

Loftus, E. F., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition*, *17*, 607–616.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 19–31.

Loftus, E. F., Schooler, J. W., & Wagenaar, W. A. (1985). The fate of memory: Comment on McCloskey and Zaragoza. *Journal of Experimental Psychology: General*, *114*, 375–387.

López, J. C. (2000). Shaky memories in indelible ink. *Nature Reviews Neuroscience*, *406*, 6.

Luttges, M. W., & McGaugh, J. L. (1967). Permanence of retrograde amnesia produced by electroconvulsive shock. *Science*, *156*, 408–410.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting interdisciplinary approaches* (pp. 1–57). Mahwah, NJ.

Mactutus, C. F., Ferek, J. M., George, C. A., & Riccio, D. C. (1982). Hypothermia-induced amnesia for newly acquired and old reactivated memories: Commonalities and distinctions. *Physiological Psychology*, *10*, 79–95.

Marin, M.-F., Pilgrim, K., & Lupien, S. J. (2010). Modulatory effects of stress on reactivated emotional memories. *Psychoneuroendocrinology*, *35*, 1388–1396.

Marsh, E. J., Meade, M. L., & Roediger, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, *49*, 519–536.

Martin, E. (1971). Verbal learning theory and independent retrieval phenomena. *Psychological Review*, *78*, 314–332.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.

McCloskey, M., & Zaragoza, M. S. (1985a). Misleading postevent information and memory for events: arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*, 1–16.

McCloskey, M., & Zaragoza, M. S. (1985b). Postevent information and memory: Reply to Loftus, Schooler, and Wagenaar. *Journal of Experimental Psychology: General*, *114*, 381–387.

McDougall, W. (1901). Experimentelle Beitrage zur Lehre vom Gedächtniss. *Mind*, *10*, 388–394.

McGaugh, J. L. (1966). Time-dependent processes in memory storage. *Science*, *153*, 1351–1358.

McGaugh, J. L. (2000). Memory–a century of consolidation. *Science*, *287*, 248–251.

McGaugh, J. L., & Dawson, R. G. (1971). Modification of memory storage processes. *Behavioral Science*, *16*, 45–63.

McGaugh, J. L., & Landfield, P. W. (1970). Delayed development of amnesia following electroconvulsive shock. *Physiology & Behavior*, *5*, 1109–1113.

McGaugh, J. L., & Petrinovich, L. F. (1966). Neural consolidation and electroconvulsive shock reexamined. *Psychological Review*, *73*, 382–387.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, *39*, 352–370.

McGeoch, J. A. (1933a). Studies in retroactive inhibition: I. the temporal course of the inhibitory effects of interpolated learning. *The Journal of General Psychology*, *9*, 24–43.

McGeoch, J. A. (1933b). Studies in retroactive inhibition: II. relationships between temporal point of interpolation, length of interval, and amount of retroactive inhibition. *The Journal* *of General Psychology*, *9*, 44–57.

McGeoch, J. A. (1942). *The psychology of human learning: An introduction*. New York: Longmans.

McGeoch, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *The American Journal of Psychology*, *43*, 579–588.

McGeoch, J. A., & McGeoch, G. O. (1937). Studies in retroactive inhibition: X. The influence of similarity of meaning between lists of paired associates. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*, 320–329.

McGeoch, J. A., & Nolen, M. E. (1933). Studies in retroactive inhibition. IV. temporal point of interpolation and degree of retroactive inhibition. *Journal of Comparative Psychology*, *15*, 407–417.

McKenzie, S., & Eichenbaum, H. (2011). Consolidation and reconsolidation: Two lives of memories? *Neuron*, *71*, 224–233.

Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*, 108–141.

Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195–244.

Meir Drexler, S., Merz, C. J., Hamacher-Dang, T. C., Marquardt, V., Fritsch, N., Otto, T., & Wolf, O. T. (2014). Effects of postretrieval-extinction learning on return of contextually controlled cued fear. *Behavioral Neuroscience*, *128*, 474–481.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*, 1–21.

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *The American Journal of Psychology*, *53*, 173–203.

Mensink, G. J., & Raaijmakers, J. G. (1988). A model for interference and forgetting.

Psychological Review, *95*, 434–455.

Milekic, M. H., & Alberini, C. M. (2002). Temporally graded requirement for protein synthesis following memory reactivation. *Neuron*, *36*, 521–525.

Miller, R. R. (1970). Effects of environmental complexity on amnesia induced by electroconvulsive shock in rats. *Journal of Comparative and Physiological Psychology*, *71*, 267–275.

Miller, R. R., & Matzel, L. D. (2000). Reconsolidation: Memory involves far more than "consolidation". *Nature Reviews Neuroscience*, *1*, 214–216.

Miller, R. R., & Matzel, L. D. (2006). Retrieval failure versus memory loss in experimental amnesia: Definitions and processes. *Learning & Memory*, *13*, 491–497.

Miller, R. R., & Oberling, P. (1998). Analogies between occasion setting and Pavlovian conditioning. In N. A. Schmajuk & P. C. Holland (Eds.), *Occasion setting associative learning and cognition in animals* (pp. 3–35).

Miller, R. R., & Springer, A. D. (1971). Temporal course of amnesia in rats after electroconvulsive shock. *Physiology & Behavior*, *6*, 229–233.

Miller, R. R., & Springer, A. D. (1972). Induced recovery of memory in rats following electroconvulsive shock. *Physiology & Behavior*, *8*, 645–651.

Miller, R. R., & Springer, A. D. (1973). Amnesia, consolidation, and retrieval. *Psychological Review*, *80*, 69–79.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*, 363–386.

Millin, P. M., Moody, E. W., & Riccio, D. C. (2001). Interpretations of retrograde amnesia: old problems redux. *Nature Reviews Neuroscience*, *2*, 68–70.

Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science*, *160*, 554–555.

Monfils, M.-H., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: key to persistent attenuation of fear memories. *Science*, *324*, 951–955.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of bayes factors for common designs. *R Package.*

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., … Zwaan, R. A. (2016). The peer reviewers' openness initiative: Incentivizing open research practices through peer review. *Royal Society Open Science*, *3*, 150547–7.

Morris, R. G. M., Inglis, J., Ainge, J. A., Olverman, H. J., Tulloch, J., Dudai, Y., & Kelly, P. A. T. (2006). Memory reconsolidation: Sensitivity of spatial memory to inhibition of protein synthesis in dorsal hippocampus during encoding and retrieval. *Neuron*, *50*, 479–489.

Morton, J., Hammersley, R. H., & Bekerian, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, *20*, 1–23.

Moscovitch, M. (2007). Why the engram is elusive. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory concepts* (pp. 17–21). Oxford University Press.

Mowrer, R. R., & Gordon, W. C. (1983). Effects of cuing in an irrelevant context. *Animal Learning & Behavior*, *11*, 401–406.

Muther, W. S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science*, *3*, 429–430.

Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift Für Psychologie Ergänzungsband*, *1*, 1–300.

Nadel, L., & Land, C. (2000). Memory traces revisited. *Nature Reviews Neuroscience*, *1*, 209–212.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*, 217–227.

Nader, K. (2003a). Memory traces unbound. *Trends in Neurosciences*, *26*, 65–72.

Nader, K. (2003b). Neuroscience: re-recording human memories. *Nature*, *425*, 571–572.

Nader, K. (2006). Unraveling the impenetrable nature of amnesia: A new beginning. *Learning & Memory*, *13*, 489–490.

Nader, K. (2015). Reconsolidation and the dynamic nature of memory. *Cold Spring Harbor Perspectives in Biology*, a021782–17.

Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews Neuroscience*, *10*, 224–234.

Nader, K., & Wang, S. H. (2006). Fading in. *Learning & Memory*, *13*, 530–535.

Nader, K., Hardt, O., Einarsson, E. Ö., & Finnie, P. S. B. (2013). The dynamic nature of memory. In C. M. Alberini (Ed.), *Memory reconsolidation* (pp. 15–41). New York: Academic Press.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000a). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*, 722–726.

Nader, K., Schafe, G. E., & LeDoux, J. E. (2000b). The labile nature of consolidation theory. *Nature Reviews Neuroscience*, *1*, 216–219.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*, 389–395.

Newton, J. M., & Wickens, D. D. (1956). Retroactive inhibition as a function of the temporal position of the interpolated learning. *Journal of Experimental Psychology: General*, *51*, 149–154.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Reviews Neuroscience*, *14*, 1105–1109.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615–631.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*, 303–330.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 1–8.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872.

Oyarzún, J. P., Lopez-Barroso, D., Fuentemilla, L., Cucurell, D., Pedraza, C., Rodriguez-Fornells, A., & Diego-Balaguer, R. de. (2012). Updating fearful memories with extinction training during reconsolidation: a human study using auditory aversive stimuli. *PLoS ONE*, *7*, 1–9.

O'Boyle, E. H., Banks, G. C., & Gonzalez-Mule, E. (2014). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*, 1–24.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*, 531–536.

Pashler, H., Kang, S. H. K., & Mozer, M. C. (2013). Reviewing erroneous information facilitates memory updating. *Cognition*, *128*, 424–430.

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*, e286.

Pavlov, I. P. (1927). *Conditioned Reflexes.* Oxford, UK: Oxford University Press.

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, *101*, 5–27.

Payne, D. G., Toglia, M. P., & Anastasi, J. S. (1994). Recognition performance level and the magnitude of the misinformation effect in eyewitness memory. *Psychonomic Bulletin & Review*, *1*, 376–382.

Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation*, *21*, 1–31.

Pedreira, M. E., Pérez-Cuesta, L. M., & Maldonado, H. (2004). Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learning & Memory*, *11*, 579–585.

Pine, A., Mendelsohn, A., & Dudai, Y. (2014). Unconscious learning of likes and dislikes is persistent, resilient, and reconsolidates. *Frontiers in Psychology*, *5*, 1–13.

Pitman, R. K. (2011). Will reconsolidation blockade offer a novel treatment for posttraumatic stress disorder? *Frontiers in Behavioral Neuroscience*, *5*, 1–2.

Poirier, M., Nairne, J. S., Morin, C., Zimmermann, F. G. S., Koutmeridou, K., & Fowler, J. (2012). Memory as discrimination: A challenge to the encoding retrieval-match principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 16–29.

Poland, G. A., & Spier, R. (2010). Fear, misinformation, and innumerates: How the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine*, *28*, 2361–2362.

Porter, P. B., & Stone, C. P. (1947). Electroconvulsive shock in rats under ether anesthesia. *Journal of Comparative and Physiological Psychology*, *40*, 441–456.

Postman, L. (1975). Verbal learning and memory. *Annual Review of Psychology*, *26*, 291–335.

Postman, L., & Alper, T. G. (1946). Retroactive inhibition as a function of the time of interpolation of the inhibitor between learning and recall. *The American Journal of Psychology*, *59*, 439–449.

Postman, L., & Stark, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology: General*, *79*, 168–177.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, *1*, 19–40.

Potts, R., & Shanks, D. R. (2012). Can testing immunize memories against interference? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1780–1785.

Power, A. E., Berlau, D. J., McGaugh, J. L., & Steward, O. (2006). Anisomycin infused into the hippocampus fails to block "reconsolidation" but impairs extinction: The role of re-exposure duration. *Learning & Memory*, *13*, 27–34.

Procter, R., Vis, F., & Voss, A. (2011, December). How riot rumours spread on Twitter. Retrieved from http://www.guardian.co.uk/

Przybyslawski, J., & Sara, S. J. (1997). Reconsolidation of memory after its reactivation. *Behavioural Brain Research*, *84*, 241–246.

Quartermain, D., Judge, M. E., & Jung, H. (1988). Amphetamine enhances retrieval following diverse sources of forgetting. *Physiology & Behavior*, *43*, 239–241.

Quartermain, D., McEwen, B. S., & Azmitia, E. C. (1970). Amnesia produced by electroconvulsive shock or cycloheximide: Conditions for recovery. *Science*, *169*, 683–686.

Quartermain, D., McEwen, B. S., & Azmitia, E. C. (1972). Recovery of memory following amnesia in the rat and mouse. *Journal of Comparative and Physiological Psychology*, *79*, 360–370.

Radyushkin, K. A., & Anokhin, K. V. (1999). Recovery of memory in chicks after disruption during learning: the reversibility of amnesia induced by protein synthesis inhibitors. *Neuroscience and Behavioral Physiology*, *29*, 31–36.

Rand, G., & Wapner, S. (1967). Postural status as a factor in memory. *Journal of Verbal Learning and Verbal Behavior*, *6*, 268–271.

Rescorla, R. A. (2001). Experimental extinction. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 119–154). Mahwah, NJ.

Rescorla, R. A. (2004). Spontaneous recovery. *Learning & Memory*, *11*, 501–509.

Rescorla, R. A., & Heth, C. D. (1975). Reinstatement of fear to an extinguished conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, *1*, 88–96.

Rescorla, R. A., & Wagner, A. D. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy

(Eds.), *Classical conditioning ii: Current research and theory* (pp. 64–99). New York: Apple-Century-Crofts.

Ribot, T. (1881). *Les maladies de la memoire*. New York: Appleton-Century-Crofts.

Ribot, T. (1883). *Diseases of memory*. New York: Fitzgerald.

Riccio, D. C., & Millin, P. M. (2003). Retrograde amnesia: Forgetting back. *Current Directions in Psychological Science*, *12*, 41–44.

Riccio, D. C., & Richardson, R. (1984). The status of memory following experimentally induced amnesias: Gone, but not forgotten. *Physiological Psychology*, *12*, 59–72.

Riccio, D. C., Millin, P. M., & Bogart, A. R. (2006). Reconsolidation: A brief history, a retrieval view, and some recent issues. *Learning & Memory*, *13*, 536–544.

Riccio, D. C., Moody, E. W., & Millin, P. M. (2002). Reconsolidation reconsidered. *Integrative Physiological & Behavioral Science*, *37*, 245–253.

Richardson, R., Riccio, D. C., & Mowrey, H. (1982). Retrograde amnesia for previously acquired Pavlovian conditioning: UCS exposure as a reactivation treatment. *Physiological Psychology*, *10*, 384–390.

Richardson, R., Riccio, D. C., Jamis, M., Cabosky, J., & Skoczen, T. (1982). Modification of reactivated memory through "counterconditioning". *The American Journal of Psychology*, *95*, 67–84.

Rickard, T. C., Cai, D. J., Rieth, C. A., Jones, J., & Ard, M. C. (2008). Sleep does not enhance motor sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 834–842.

Robbins, S. J. (1990). Mechanisms underlying spontaneous recovery in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*, 235–249.

Robinson, E. S. (1920). Some factors determining the degree of retroactive inhibition, *28*, 1–57.

Rodriguez, W. A., Rodriguez, S. B., Phillips, M. Y., & Martinez, J. L. (1993). Post-reactivation cocaine administration facilitates later acquisition of an avoidance response in rats. *Behavioural Brain Research*, *59*, 125–129.

Rodriguez-Ortiz, C. J. (2005). Protein synthesis underlies post-retrieval memory consolidation to a restricted degree only when updated information is obtained. *Learning & Memory*, *12*, 533–537.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27.

Roediger, H. L., & Guynn, M. J. (1996). Retrieval processes. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 197–236). London: Academic Press.

Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1155–1159.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: General*, *21*, 803–814.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.

Rosenthal, R. (1990). Replication in behavioral research. *Journal of Social Behavior & Personality*, *5*, 1–30.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237.

Roullet, P., & Sara, S. J. (1998). Consolidation of memory after its reactivation: Involvement of ß noradrenergic receptors in the late phase. *Neural Plasticity*, *6*, 63–68.

Russell, W. R., & Nathan, P. W. (1946). Traumatic amnesia. *Brain: A Journal of Neurology*, *69*, 280–300.

Sangha, S., Scheibenstock, A., & Lukowiak, K. (2003). Reconsolidation of a long-term memory in Lymnaea requires new protein and RNA synthesis and the soma of right pedal dorsal 1. *The Journal of Neuroscience*, *23*, 8034–8040.

Sara, S. J. (1973). Recovery from hypoxia and ECS-induced amnesia after a single exposure to training environment. *Physiology & Behavior*, *10*, 85–89.

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

Sara, S. J. (2000b). Strengthening the shaky trace through retrieval. *Nature Reviews Neuroscience*, *1*, 212–213.

Sara, S. J., & Hars, B. (2006). In memory of consolidation. *Learning & Memory*, *13*, 515–521.

Sara, S. J., & Remacle, J. F. (1977). Strychnine-induced passive avoidance facilitation after electroconvulsive shock or undertraining: a retrieval effect. *Behavioral Biology*, *19*, 465–475.

Schacter, D. L., Eich, J. E., & Tulving, E. (1978). Richard Semon's theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *17*, 721–743.

Schafe, G. E., & LeDoux, J. E. (2000). Memory consolidation of auditory pavlovian fear conditioning requires protein synthesis and protein kinase A in the amygdala. *The Journal of Neuroscience*, *20*, 1–5.

Schiller, D., & Phelps, E. A. (2011). Does reconsolidation occur in humans? *Frontiers in Behavioral Neuroscience*, *5*, 1–12.

Schiller, D., Kanen, J. W., LeDoux, J. E., Monfils, M.-H., & Phelps, E. A. (2013). Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. *Proceedings of the National Academy of Sciences of The United States of America*, *110*, 20040–20045.

Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms.

Nature, *463*, 49–53.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*, 90–100.

Schneider, A. M., & Sherman, W. (1968). Amnesia: A function of the temporal relation of footshock to electroconvulsive shock. *Science*, *159*, 219–221.

Schwabe, L., & Wolf, O. T. (2009). New episodic learning interferes with the reconsolidation of autobiographical memories. *PLoS ONE*, *4*, 1–4.

Schwabe, L., & Wolf, O. T. (2010). Stress impairs the reconsolidation of autobiographical memories. *Neurobiology of Learning and Memory*, *94*, 153–157.

Schwabe, L., Bohringer, A., Chatterjee, M., & Schachinger, H. (2008). Effects of pre-learning stress on memory for neutral, positive and negative words: Different roles of cortisol and autonomic arousal. *Neurobiology of Learning and Memory*, *90*, 44–53.

Schwabe, L., Nader, K., & Pruessner, J. C. (2013). β-Adrenergic blockade during reactivation reduces the subjective feeling of remembering associated with emotional episodic memories. *Biological Psychology*, *92*, 227–232.

Schwabe, L., Nader, K., & Pruessner, J. C. (2014). Reconsolidation of human memory: Brain mechanisms and clinical relevance. *Biological Psychiatry*, *76*, 274–280.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review*, *18*, 455–468.

Seifert, C. M. (2002). The continued influence of misinformation in memory: What makes a correction effective? *Psychology of Learning and Motivation*, *41*, 265–292.

Semon, R. (1921). The Mneme. London: George Allen & Unwin.

Serota, R. G. (1971). Acetoxycycloheximide and transient amnesia in the rat. *Proceedings of the National Academy of Sciences of The United States of America*, *68*, 1249–1250.

Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learning & Memory*, *21*, 580–584.

Silverstein, A. (1967). Unlearning, spontaneous recovery, and the partial-reinforcement effect in paired-associate learning. *Journal of Experimental Psychology: General*, *73*, 15–21.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76–80.

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to Registered Replication Reports at Perspectives on Psychological Science. *Perspectives on Psychological Science*, *9*, 552–555.

Sisson, E. D. (1939). Retroactive inhibition: The temporal position of interpolated activity. *Journal of Experimental Psychology: General*, *25*, 228–233.

Skaggs, E. B. (1933). A discussion of the temporal point of interpolation and degree of retroactive inhibition. *Journal of Comparative Psychology*, *16*, 411–414.

Slamecka, N. J. (1966). A search for spontaneous recovery of verbal associations. *Journal of Verbal Learning and Verbal Behavior*, *5*, 205–207.

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1712–1725.

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

Soeter, M., & Kindt, M. (2010). Dissociating response systems: Erasing fear from memory. *Neurobiology of Learning and Memory*, *94*, 30–41.

Soeter, M., & Kindt, M. (2011). Disrupting reconsolidation: Pharmacological and behavioral manipulations. *Learning & Memory*, *18*, 357–366.

Spear, N. E. (1973). Retrieval of memory in animals. *Psychological Review*, *80*, 163–194.

Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General*, *139*, 266–298.

Spring, J. D., Wood, N. E., Pfeiffer, C. M., Milad, M. R., Pitman, R. K., & Orr, S. P. (2015). Prereactivation propranolol fails to reduce skin conductance reactivity to prepared fear-conditioned stimuli. *The Psychology of Learning and Motivation*, *52*, 407–415.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*, 195–231.

Squire, L. R. (2006). Lost forever or temporarily misplaced? The long debate about the nature of memory impairment. *Learning & Memory*, *13*, 522–529.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology*, *5*, 169–177.

Squire, L. R., & Barondes, S. H. (1972). Variable decay of memory and its recovery in cycloheximide-treated mice. *Proceedings of the National Academy of Sciences of The United States of America*, *69*, 1416–1420.

Squire, L. R., Slater, P. C., & Chace, P. M. (1975). Retrograde amnesia: temporal gradient in very long term memory following electroconvulsive therapy. *Science*, *187*, 77–79.

Squire, L. R., Slater, P. C., & Chace, P. M. (1976). Reactivation of recent or remote memory before electroconvulsive therapy does not produce retrograde amnesia. *Behavioral Biology*, *18*, 335–343.

Squire, L. R., Slater, P. C., & Miller, P. L. (1981). Retrograde amnesia and bilateral electroconvulsive therapy. Long-term follow-up. *Archives of General Psychiatry*, *38*, 89–95.

St Jacques, P. L., & Schacter, D. L. (2013). Modifying memory: Selectively enhancing and updating personal memories for a museum tour by reactivating them. *Psychological Science*, *24*, 537–543.

St Jacques, P. L., Montgomery, D., & Schacter, D. L. (2015). Modifying memory for a museum tour in older adults: Reactivation-related updating that enhances and distorts memory is reduced in ageing. *Memory*, *23*, 876–887.

St Jacques, P. L., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences of The United States of America*, *110*, 19671–19678.

Steinfurth, E. C. K., Kanen, J. W., Raio, C. M., Clem, R. L., Huganir, R. L., & Phelps, E. A. (2014). Young and old Pavlovian fear memories can be modified with extinction training during reconsolidation in humans. *Learning & Memory*, *21*, 338–341.

Sterne, J., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*, 1046–1055.

Stollhoff, N. (2005). Spontaneous recovery from extinction depends on the reconsolidation of the acquisition memory in an appetitive learning paradigm in the honeybee (apis mellifera). *The Journal of Neuroscience*, *25*, 4485–4492.

Strange, B. A., Kroes, M. C. W., Fan, J. E., & Dolan, R. J. (2010). Emotion causes targeted forgetting of established memories. *Frontiers in Behavioral Neuroscience*, *4*, 1–13.

Suzuki, A. (2004). Memory reconsolidation and extinction have distinct temporal and biochemical signatures. *The Journal of Neuroscience*, *24*, 4787–4795.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1392–1399.

Takarangi, M. K. T., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: the development of a new stimulus set. *Applied Cognitive Psychology*, *20*, 583–590.

Taubenfeld, S. M., Milekic, M. H., & Monti, B. (2001). The consolidation of new but not reactivated memory requires hippocampal C/EBPβ. *Nature Neuroscience*, *4*, 813–818.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013).

General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods*, *45*, 1115–1143.

Thompson, C. I., & Neely, J. E. (1970). Dissociated learning in rats produced by electroconvulsive shock. *Physiology & Behavior*, *5*, 783–786.

Thune, L. E., & Underwood, B. J. (1943). Retroactive inhibition as a function of degree of interpolated learning. *Journal of Experimental Psychology: General*, *32*, 185–200.

Titcomb, A. L., & Reyna, V. F. (1995). Memory interference and misinformation effects. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 263–294). New York: Elsevier.

Toglia, M. P., Ross, D. F., Ceci, S. J., & Hembrooke, H. (1992). The suggestibility of children's memory: A social-psychological and cognitive interpretation. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 217–237). New York, NY: Springer New York.

Trent, S., Barnes, P., Hall, J., & Thomas, K. L. (2015). Rescue of long-term memory after reconsolidation blockade. *Nature Communications*, *6*, 1–7.

Tronson, N. C., & Taylor, J. R. (2007). Molecular mechanisms of memory reconsolidation. *Nature Reviews Neuroscience*, *8*, 262–275.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*, 175–184.

Tulving, E. (1974). Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *American Scientist*, *62*, 74–82.

Tulving, E. (1983a). Ecphoric processes in episodic memory. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences.*, *302*, 361–371.

Tulving, E. (1983b). *Elements of Episodic Memory.* New York: Oxford University Press.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of*

Psychology, *21*, 437–484.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381–391.

Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology: General*, *87*, 1–8.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

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

Tulving, E., & Watkins, M. J. (1975). Structure of memory traces. *Psychological Review*, *82*, 261–275.

Tversky, B., & Tuchin, M. (1989). A reconciliation of the evidence on eyewitness testimony: Comments on McCloskey and Zaragoza. *Journal of Experimental Psychology: General*, *118*, 86–91.

Underwood, B. J. (1948). "Spontaneous recovery" of verbal associations. *Journal of Experimental Psychology: General*, *38*, 429–439.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76*, 559–573.

Van der Loo, M. (2014). The stringdist package for approximate string matching. *The R Journal*, *6*, 111–122.

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., & Storms, G. (2015). Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra*, *1*, 1–5.

Verne, J. (1996). *Journey to the centre of the earth*. Ware: Wordsworth Editions Limited.

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: State of the art. *Annual Review of Clinical Psychology*, *9*, 215–248.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Maas, H. L. J. van der, & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*, 632–638.

Walker, M. P., & Stickgold, R. (2016). Understanding the boundary conditions of memory reconsolidation. *Proceedings of the National Academy of Sciences of The United States of America*, *113*, E3991–2.

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, *425*, 616–620.

Warren, V. T., Anderson, K. M., Kwon, C., Bosshardt, L., Jovanovic, T., Bradley, B., & Norrholm, S. D. (2014). Human fear extinction and return of fear using reconsolidation update mechanisms: The contribution of on-line expectancy ratings. *Neurobiology of Learning and Memory*, *113*, 165–173.

Watkins, M. J. (1990). Mediationism and the obfuscation of memory. *American Psychologist*, *45*, 328–335.

Watkins, M. J., & Watkins, O. C. (1976). Cue-overload theory and the method of interpolated attributes. *Bulletin of the Psychonomic Society*, *7*, 289–291.

Wheeler, M. A. (1995). Improvement in recall over time without repeated testing: Spontaneous recovery revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 173–184.

Wichert, S., Wolf, O. T., & Schwabe, L. (2011). Reactivation, interference, and reconsolidation: Are recent and remote memories likewise susceptible? *Behavioral Neuroscience*, *125*, 699–704.

Wichert, S., Wolf, O. T., & Schwabe, L. (2013a). Changing memories after reactivation: A one-time opportunity? *Neurobiology of Learning and Memory*, *99*, 38–49.

Wichert, S., Wolf, O. T., & Schwabe, L. (2013b). Updating of episodic memories depends on the strength of new learning after memory reactivation. *Behavioural Neuroscience*, *127*, 331–338.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726–728.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *40*, 361–387.

Wilkes, A. L., & Reynolds, D. J. (1999). On certain limitations accompanying readers' interpretations of corrections in episodic text. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *52*, 165–183.

Wilkinson, D., & Halligan, P. (2004). The relevance of behavioural measures for functional-imaging studies of cognition. *Nature Reviews Neuroscience*, *5*, 67–73.

Windschitl, P. D. (1996). Memory for faces: Evidence of retrieval-based impairment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1101–1122.

Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, *17*, 766–780.

Wirkner, J., Löw, A., Hamm, A. O., & Weymar, M. (2015). New learning following reactivation in the human brain: Targeting emotional memories through rapid serial visual presentation. *Neurobiology of Learning and Memory*, *119*, 63–68.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*, 235–269.

Xue, Y. X., Luo, Y. X., Wu, P., Shi, H. S., Xue, L. F., Chen, C., ... Lu, L. (2012). A memory retrieval-extinction procedure to prevent drug craving and relapse. *Science*, *336*, 241–245.

Yacoby, A., Dudai, Y., & Mendelsohn, A. (2015). Metamemory ratings predict long-term changes in reactivated episodic memories. *Frontiers in Behavioral Neuroscience*, *9*, e20.

Young, A. G., & Galluscio, E. H. (1971). Recovery from ECS-produced amnesia. *Psychonomic Science*, *22*, 149–151.

Zaragoza, M. S., & Koshmider, J. W. (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 246–255.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 934–945.

Zaragoza, M. S., & McCloskey, M. (1989). Misleading postevent information and the memory impairment hypothesis: Comment on Belli and reply to Tversky and Tuchin. *Journal of Experimental Psychology: General*, *118*, 92–99.

Zaragoza, M. S., Belli, R. F., & Payment, K. E. (2006). Misinformation effects and the suggestibility of eyewitness memory. In M. Garry & H. Hayne (Eds.), *Do justice and let the sky fall: Elizabeth f. loftus and her contributions to science, law, and academic freedom* (pp. 35–63). London: Lawrence Erlbaum Associates.

Zaragoza, M. S., Dahlgren, D., & Muench, J. (1992). The role of memory impairment in children's suggestibility. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 184–216). New York, NY: Springer New York.

Zaragoza, M. S., McCloskey, M., & Jamis, M. (1987). Misleading postevent information and recall of the original event: Further evidence against the memory impairment hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 36–44.

Zhao, L.-Y., Zhang, X.-L., Shi, J., Epstein, D. H., & Lu, L. (2009). Psychosocial stress after reactivation of drug-related memory impairs later recall in abstinent heroin addicts. *Psychopharmacology*, *203*, 599–608.

Zinkin, S., & Miller, A. J. (1967). Recovery of memory after amnesia induced by electroconvulsive shock. *Science*, *155*, 102–104.