Forgetting in memory models:
Arguments against trace decay and consolidation failure

Gordon D. A. Brown

University of Warwick and University of Western Australia

Stephan Lewandowsky

University of Western Australia

1. Introduction

The chapters in this collection all reference causes of forgetting. But the variety of possible causes (and continued lack of consensus regarding them in different strands of the literature) is striking. Here we examine insights into trace decay, interference, and consolidation that have emerged from recent computational and mathematical models of memory. We suggest that such models (a) allow rejection of temporal decay as a primary cause of forgetting even in short-term memory tasks; (b) undermine the inference from forgetting data to a distinction between separate short-term and long-term memory systems (STS vs. LTS), and (c) offer an alternative explanation, in terms of temporal distinctiveness and interference, for most if not all of the behavioural evidence that has previously been taken as evidence for consolidation.

Two key theoretical issues underpin the present discussion. The first of these concerns the putative distinction between two memory systems that are dedicated to the storage of information over the short and the long term (STS and LTS respectively), and the second concerns the importance of consolidation failure as a cause of forgetting.

STS vs. LTS. Although the utility of a theoretical distinction between STS and LTS has often been questioned (Crowder, 1989; Melton, 1963) only recently have specific models emerged that claim to account for both short-term and long-term memory phenomena within a unified framework. For example, one of our own models (G. D. A. Brown, Neath, & Chater, 2007) asserts that the mechanisms underlying retrieval and forgetting are the same over both short and long timescales, thus questioning the case for a STS-LTS distinction. Here we focus on just one of the traditional arguments for a distinction between memory systems– viz. the assumption of different causes of forgetting from STS (temporal decay) and LTS (proactive and retroactive interference). To anticipate: we conclude that there is no evidence that time-based decay is the sole or even primary cause of forgetting over the short term, thus undermining one piece of evidence from forgetting for the traditional distinction.

Other arguments that are consistent with our perspective can be found elsewhere (e.g. Bhatarah, Ward, & Tan, 2008; G. D. A. Brown, Chater, & Neath, 2008; G. D. A. Brown, Della Sala, Foster, & Vousden, 2007; Neath & Brown, 2006; Tan & Ward, 2000) (although see Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005). In particular, we note that interference-based models such as those of Lewandowsky and others (Lewandowsky & Farrell, 2008; Oberauer & Lewandowsky, 2008) and Brown et al. (2007) can account for forgetting data that have previously been assumed to implicate a STS-LTS distinction, and that empirical evidence that has been taken in support of temporal decay can be reinterpreted (Lewandowsky & Brown, 2005; Lewandowsky & Oberauer, 2008; Lewandowsky, Oberauer, & Brown, 2009a).

Consolidation. The second theoretical issue addressed by the present chapter is that of consolidation as a primary factor underpinning memory and forgetting. Consolidation refers to the idea that memories continue to strengthen *after* they have been formed, and that they thus become more resilient to forgetting over time (e.g. Wixted, 2005). As Wixted (2004b, 2005) notes, consolidation has featured prominently in theorizing on forgetting in the neurosciences for several decades, whereas most cognitive approaches have relied exclusively on alternative notions such as interference or decay. Indeed, not a single recent formal model of memory within a cognitive tradition ascribes an important role to consolidation (e.g. Botvinick & Plaut, 2006; G. D. A. Brown, Neath, et al., 2007; G. D. A. Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999, 2006; Davelaar, et al., 2005; Gillund & Shiffrin, 1984; Henson, 1998; Howard & Kahana, 2002; Oberauer & Lewandowsky, 2008; Page & Norris, 1998; Polyn, Norman, & Kahana, 2009; Raaijmakers & Shiffrin, 1981; Sederberg, Howard, & Kahana, 2008; Shiffrin & Steyvers, 1997). Although these models differ in many respects (e.g. whether forgetting is due primarily to temporal decay, or interference, or encoding failure, or lack of temporal distinctiveness) there is remarkable, albeit implicit, agreement about the unimportance of consolidation. Notwithstanding their omission of consolidation processes, these models can explain a plethora of findings, including a variety of forgetting phenomena. Perhaps not surprisingly, an isolated but notable exception to the models' success is their inability to address data that have been taken as evidence for the importance of consolidation (see M. T. Dewar, Cowan, & Della Sala, 2007).

Conversely, models of memory that do emphasize consolidation (McClelland, McNaughton, & O'Reilly, 1995; Meeter & Murre, 2005; Norman & O'Reilly, 2003) handle the data implicating consolidation, but they typically do not address the rich data sets that are traditionally taken as the explananda for cognitive models of memory.

The central question we address, then, is the following: Are current cognitive models deficient because they include no provision for consolidation? Or might the behavioural evidence that has been taken as strong support for consolidation be interpretable in other ways? To foreshadow our principal conclusion, we endorse the latter hypothesis by suggesting that a temporal distinctiveness model of memory can account for data that have hitherto been taken to implicate consolidation.

We proceed as follows. First, we review and reinterpret several sources of evidence for temporal decay, and we conclude that decay is not a primary cause of forgetting in the short term. Second, we consider the form of the forgetting function, and show that (contrary to previous views) it provides support neither for a distinction between STS and LTS nor (as has sometimes been argued; see below) for consolidation. Third, we explore well-known cases where memory *improves* over time (the recency-to-primacy shift) and show that such phenomena not only provide further evidence against trace decay but can also be readily interpreted without recourse to consolidation. In the fourth and final section of the chapter, we directly confront evidence for consolidation in memory and reinterpret the data within a cognitive model that includes no consolidation. Note that owing to space constraints, we restrict consideration to human *behavioural* data only; we do not consider data from imaging studies or lesioning studies involving nonhuman animals. We recognize the potential importance of those sources of data but they are beyond the scope of the present chapter; a case for consolidation as a psychological variable must in any case be supported by behavioural data.

2. Reinterpreting evidence against forgetting due to time-based decay

There has been a long-standing consensus that decay plays no role in forgetting over the long term (Jenkins & Dallenbach, 1924). Scholars of short-term memory likewise initially eschewed the notion of decay (Atkinson & Shiffrin, 1971), although it gained prominence with Baddeley's phonological loop model (Baddeley, Thomson, & Buchanan, 1975) and decay continues to be central to a number of recent models of short-term memory (e.g. Burgess & Hitch, 1999; Page & Norris, 1998). This theoretical commitment to short-term decay sits alongside a pervasive agreement in the field that long-term forgetting does not involve temporal decay. Parsimony alone implies that a unitary time-invariant forgetting mechanism would be preferable, and we now show that this preference is buttressed by much empirical support.

What empirical evidence could differentiate between time-based decay and other forms of forgetting? At first glance, this issue may appear trivial: One simply extends the amount of time that information resides in memory and observes how much additional forgetting occurs. Alas, closer inspection reveals two problems that render the issue far from trivial. First, it is a priori unclear exactly how much forgetting would be expected on a decay view. Suppose recall declines from .81 to .80 after a few seconds' delay; is this evidence for decay? What about a decline from .80 to .50? This problem is best resolved by interpreting data only with respect to *quantitative* predictions of the models under consideration (Lewandowsky, et al., 2009a; Oberauer & Lewandowsky, 2008). Second, when confronted with unanticipated outcomes, theorists can invoke auxiliary processes to explain the data. For example, if forgetting is absent, decay theorists can appeal to surreptitious compensatory rehearsal that reverses the effects of decay, thus masking its presence (Vallar & Baddeley, 1982). Conversely, an interference view can handle unpredicted forgetting by postulating that some activity during retention interfered with memory (Lewandowsky, Geiger, & Oberauer, 2008). In order to avoid these interpretative problems, two conditions must be met: (a) rehearsal must be controlled and (b) retention intervals must be kept free of interference. Unfortunately, these two goals are in conflict with each other. To disable rehearsal, there must be some cognitive activity (e.g., the overt recitation of irrelevant material), but this activity could also create interference. Two methodologies have recently emerged that satisfy both goals.

Berman, Jonides, and Lewis (2009) exploited the fact that in short-term recognition, negative probes that were on the preceding trial's study list are generally rejected more slowly than completely novel lures. For example, response latencies to the probe "lion" after study of the list "cat", "table", "truck" are dependent on whether or not "lion" had been studied on the preceding trial—notwithstanding the fact that the preceding list is now irrelevant rehearsal of those items entirely counter-productive. Berman et al. found that this disadvantage for recently-studied lures diminished only negligibly when the inter-trial interval was increased from .3 to 10 s, whereas it was eliminated by insertion of a single intervening study-test trial of equal (10 s) duration. Thus, contrary to what would be expected on a decay view, no-longer relevant information lingers in short-term memory undiminished over time unless cleared by intervening cognitive events. This methodology satisfies the two constraints just mentioned because (a) rehearsal of no-longer relevant material after its test is counter-productive and hence assuming its presence is difficult to justify, and (b) inter-trial intervals were entirely free of interfering activity.

The second methodology was developed by Lewandowsky, Duncan, and Brown (2004) and involved blocking of rehearsal during immediate serial recall by overt articulation of an irrelevant word. Retention time was varied by training participants to recall at different speeds (.4, .8, and 1.6 s/item), thus delaying recall of the last item by over 5 s at the slowest compared to the fastest speed. This added delay reduced performance only negligibly, suggesting that although compensatory rehearsal was blocked by articulation, there was little manifestation of temporal decay. A similar result was obtained by Cowan et al. (2006) using a related procedure. Children were either asked to recall a list at "whatever speed seemed best" or "as quickly as possible." Recall times decreased from .82 s/item to .5 s/item (a speed-up in excess of 30%) but left recall essentially unchanged. Although rehearsal was not explicitly controlled, it appears unlikely that when instructed to recall at a comfortable pace, children would have withheld their responses merely to rehearse.

In a recent extension to this methodology, Oberauer and Lewandowsky (2008) added yet another task during retrieval to block possible "attentional" forms of refreshing (e.g. Hudjetz & Oberauer, 2007) that might augment conventional articulatory rehearsal. In addition to overt articulation of an irrelevant word, participants performed a symbolic two-alternative choice task in between recalling list items. Increasing the number of articulations and choice responses from 1 to 4 significantly delayed recall (by up to 14 s for the last list item) but had only a negligible effect on memory (reducing accuracy by .005 per second additional delay). This appears to be a general result, holding across many experiments (Lewandowsky, et al., 2009a). Importantly, when all experiments are considered together, the data exhibit considerably less forgetting than the minimum amount that decay models must predict (Lewandowsky, et al., 2009a). Intriguingly, these findings mesh well with recent studies of forgetting in amnesia, which also find that memory over intermediate time periods can be substantially improved under conditions where interference is minimised (Cowan, Beschin, & Della Sala, 2004; Della Sala, Cowan, Beschin, & Perini, 2005).

To explain the results from the studies just reviewed, proponents of inexorable temporal decay would need to argue that some form of memory refreshing persisted despite a variety of measures to the contrary. Specifically, decay proponents would need to argue that people continued to rehearse tested and no-longer-relevant material (Berman, et al., 2009); they would have to assume that people were rehearsing while articulating irrelevant material out loud (Lewandowsky, et al., 2004); and they would have to assume that people were rehearsing while articulating out loud *and* performing an attention-demanding symbolic choice task (Oberauer & Lewandowsky, 2008). We do not consider those arguments plausible.

How can these recent data that provide evidence against decay in immediate memory be reconciled with the pervasive "word-length effect" (WLE)? The WLE refers to the finding that words that take longer to pronounce (e.g., "hippopotamus", "confederacy") are sometimes remembered more poorly than short words ("buck", "pink"). On a decay hypothesis, the fact that differences in pronunciation durations of only 150-200 ms per word result in poorer recall arises because long words have more time to decay before they can be rehearsed or output. At first glance, the WLE strongly implicates decay and it has been cited as "... perhaps the best remaining solid evidence in favour of temporary memory storage" (Cowan, 1995, p.42).

We offer a different perspective based on arguments recently advanced by Lewandowsky and Oberauer (2008). Their principal argument rests on the fact that the WLE represents a correlation between two measures—articulation duration and memory—and it therefore inherits all interpretative problems that beset correlations. The WLE is correlational irrespective of whether one compares words of different syllabic complexity ("hippopotamus" vs. "gun") or restricts consideration to a purely duration-based WLE involving words of equal syllabic complexity but differing pronunciation durations (e.g., "platoon'" vs. "racket'"). Articulation duration is, in principle, correlated with many other features that influence a word's memorability. Hence, articulation duration may simply be a proxy variable for something else that determines memorability, and notwithstanding commendably thorough attempts to the contrary (Mueller, Seymour, Kieras, & Meyer, 2003), it is impossible in principle to identify, let alone control, all of these correlated features. This lack of control opens the door for alternative explanations of the WLE not involving decay. Accordingly, a number of studies have found that the WLE arises only with some particular stimuli (Baddeley, et al., 1975), is absent in others (Lovatt, Avons, & Masterson, 2000), and sometimes even reversed (i.e., longer words are recalled better) for yet other stimuli (Caplan, Rochon, & Waters, 1992; Lovatt, et al., 2000; Lovatt, Avons, & Masterson, 2002; Neath, Bireta, & Surprenant, 2003).[1]

Another recent line of research that has been taken as evidence for trace decay comes from studies of the time-based resource-sharing (TBRS) model (Barrouillet, Bernardin, & Camos, 2004; Barrouillet & Camos, 2009). For example, using a complex span task, Portrat, Barrouillet, and Camos (2008) found that memory is reduced when the time taken to complete a between-item processing task is increased while the time available for rehearsal is controlled. Such results have been taken to reflect the operation of time-based decay. However, Lewandowsky and Oberauer (in press) showed instead that the Portrat et al. results were due to attentional processes occurring after errors on the distractor task, and Lewandowsky, Oberauer, and Brown (2009b) argued more generally against the suggestion that complex-span data implicate decay.

We conclude that a strong case can be made against a role for trace decay in forgetting over the short term.

3. The form of the forgetting curve

If our preceding arguments against decay are correct, it follows that differential causes of forgetting cannot be used to motivate a distinction between separate STS and LTS systems. However, irrespective of any appeal to short-term decay, forgetting data

[1] Lewandowsky and Oberauer's [2008] analysis is more extensive, but a full consideration of their arguments is beyond the scope of this chapter.

have been used to argue for a distinction between two memory systems in a different way, based purely on the putative form of the forgetting function. Specifically, it has been argued that the form of the forgetting curve is different over the short-term and the long-term, consistent with the suggestion that different memory systems are involved at different time scales. In addition, the form of the forgetting curve has been used as evidence for consolidation. Specifically, the fact that when two memories are of equal strength, forgetting appears to be slower for the older of the two (Jost's Second Law: Wixted, 2004a) has been taken to support the idea that older memories are more resistant to forgetting because they have had more time to consolidate (see e.g. Wixted, 2004a; 2004b for a critique of alternative interpretations). In the following, we argue against both of these inferences that have been drawn from the form of the forgetting function.

We begin by highlighting the fact that relatively few recent formal models of memory have been applied to the form of forgetting functions. Although there have been sophisticated and extended *empirical* attempts to determine the equation that best characterises the forgetting function (e.g. Rubin & Wenzel, 1994; Wixted & Ebbesen, 1990, 1997), a parallel effort involving process-level or mathematical models has been largely absent (although see Anderson & Schooler, 1991; Sikstrom, 2002). As a first step towards redressing this deficiency, we present a simple temporal distinctiveness model which assumes a single mechanism for forgetting at both short and long timescales. In the model, all forgetting is due to interference and there is no role for consolidation—however, we show that the model nonetheless (a) gives rise to forgetting functions of a variety of forms and (b) handles Jost's second law. To foreshadow, we conclude (a) that the search for "the" form of the forgetting function will necessarily remain inconclusive and (b) that the generally agreed characteristics of the forgetting curve – viz. that older memories are forgotten more slowly; Jost's Second Law – does not implicate consolidation.

What are the relevant properties of the forgetting curve? A number of researchers have suggested that the time course of forgetting is well described by a power law (e.g. Anderson & Schooler, 1991; Wixted & Ebbesen, 1990, 1997), while others have argued that it is not (Chechile, 2006; Rubin, Hinton, & Wenzel, 1999; Wickens, 1999). A power function has the form $P = aT^{-b}$ where P is the measure of memory performance; T is time elapsed, and *a* and *b* are constants. In contrast to the exponential function ($P = a\ e^{-bT}$) which is characterized by a constant rate of loss, the power function shows initially-rapid forgetting that then slows over time. This is a desirable property because ongoing debate notwithstanding, it is widely accepted that the rate of forgetting slows over time. If you can remember 100 French vocabulary words from your school education 20 years ago, and I can remember 100 French vocabulary words from 200 I learnt yesterday, it seems likely that you will remember more words than I do in a week's time (assuming that neither of us engages in any further learning in the meantime).

We (G. D. A. Brown, Neath, et al., 2007) used a simple temporal distinctiveness model, SIMPLE (for Scale Invariant Memory, Perception, and LEarning) to address the relationship between model architecture and the form of the forgetting function over different timescales. Informally, SIMPLE assumes that the confusability between any two memory traces depends on the ratio of the times that have elapsed between their encodings and the time of recall. The lower that ratio, the lower the confusability among items and hence the more likely it is that an item is recalled correctly. Hence recent items are less confusable and hence more memorable than are more distant events. For example, items that were encoded 1 s and 2 s ago are less confusable (ratio of .5) than are items from 5 and 6 seconds ago (.83). The mechanism also favours items that were separated in time over others that occurred in close succession. For example, items that occurred 5 s and 10 s ago (ratio .5) are less confusable than items that occurred 7 s and 8 s ago (.88), even though the average retention interval is equal for both pairs of items. It follows that items from further in the past, and items that occurred near each other in time, will be more difficult to recall. As we will see below, this property causes the model to expect slower forgetting of older items without recourse to consolidation.[2]

More formally, the three key assumptions of the model are as follows. (1) Items are represented by their position within a multi-dimensional psychological space, with

[2] The model as just described predicts time-based forgetting in much the same manner as a decay model and is thus challenged by the data reviewed earlier. Those challenges are overcome by modifications to the model that are not relevant to the current argument; (for details, see Lewandowsky, et al., 2004).

one of those dimensions necessarily devoted to representing time. In the present treatment we would be concerned only with this temporal dimension. (2) The similarity between any two items in memory is a declining function of the distance separating them in psychological space. (3) The probability of recalling an item is inversely proportional to that item's summed similarity to all other response alternatives, as illustrated above by the ratios between elapsed times of item pairs.

These assumptions are implemented in the model as follows (the following section may be omitted for readers not interested in the technical details). A more complete specification of the model, including its application to multidimensional memory representations, can be found in Brown, Neath, et al., (2007).

Encoding in multi-dimensional space. Memory representations are organized along a temporal dimension that reflects the (logarithmically transformed) time since their encoding.

Similarity-distance metric. Following the categorization literature, SIMPLE assumes that the similarity of any two items in memory is a reducing exponential function of the distance between them in psychological space:

$$\eta_{i,j} = \mathrm{e}^{-cd_{i,j}},$$

where $\eta_{i,j}$ is the similarity between items i and j and $d_{i,j}$ the distance between them (i.e., in this instance, the distance along the temporal axis that separates the two items). Because the time scale is assumed to be logarithmically transformed, the similarity between two items that are differentiated only along the temporal dimension can be equivalently expressed as the ratio of their temporal distances raised to the power c, thus permitting the intuitive analysis presented earlier.

Items that are very close have a similarity approaching unity (i.e., their distance is near 0 and hence the ratio of their temporal distances will be close to 1.0), whereas items that are more psychologically distant have a similarity that, in the extreme, approaches zero. The parameter c governs the rate of decline of similarity with distance. When combined with the logarithmic transformation of the temporal dimension, this similarity metric gives rise to the distinctiveness ratios mentioned earlier.

Similarity determines recall. The distinctiveness and hence discriminability of item i is inversely proportional to its summed similarity to every other potentially recallable item. Specifically, the discriminability of the memory trace for item i, D_i, is given by:

$$D_i = \frac{1}{\sum_{k=1}^{n}(\eta_{i,k})},$$

where n is the number of available response alternatives (normally this is just the number of list items).

In the full version of the model, discriminability translates into recall probability by taking into account the possibility of omissions. Omissions arise from thresholding of low retrieval probabilities by a sigmoid function: If D_i is the discriminability given by the preceding equation, the recall probability P_i is derived as:

$$P_i = \frac{1}{1+e^{-s(D_i - t)}},$$

where t is the threshold and s determines the slope (or noisiness) of the transforming function. Any D_i that falls below the threshold engenders an omission.

We use the basic assumptions just described to examine forgetting as a function of time in the model. Forgetting in the model occurs over time not because of decay but because of interference –memories become less distinguishable from one another, and hence harder to retrieve, as they retreat into the temporal distance and lose temporal distinctiveness. A more complete account of the model's forgetting behaviour is given in Brown, Neath, et al., (2007); here we merely summarise findings of theoretical interest in the current context.

First of all, we found that small alterations in the parameters of the model, which were intuitively insignificant in terms of its underlying architecture, could change the apparent form of forgetting curve produced. For example, the forgetting curve might be better described by a logarithmic function than a power law for some parameter settings, while the reverse could hold under different parameter settings (detailed model comparison to take into account the different flexibility of different functional forms was not undertaken however). This sensitivity to parameter values was taken to suggest that there need be no simple correspondence between a model architecture and the form of the forgetting function that it predicts. It may therefore be that more than a century of effort (Rubin & Wenzel, 1994) of attempting definitively to establish the form of forgetting curve (without a universally accepted result as yet) may have been misguided in the sense that the inference from the form of a forgetting curve to model architecture, or vice versa, may be far from transparent (see also Wickens, 1999).

Second, despite the above, we found that the model generally obeyed Jost's Second Law in that forgetting slowed over time. In other words, under a range of parameter settings, older memories were forgotten more slowly than younger memories of the same strength. A typical forgetting curve is shown in Figure 1 (described below). Reduced rates of forgetting over time occur naturally as a result of the ratio properties of the model. The confusability and hence discriminability in memory of any two items is, in the model, dependent on the ratio of the temporal distances of those items. That ratio will gradually approach unity as the items received into the past, but the rate at which this happens will slow down over time. For example, consider two items that occurred 1 s ago and 2 s ago. Their confusability is 1/2. When a further 1 s of retention interval was passed, their confusability will have increased to 2/3. Now consider in contrast two items that occurred 100 s and 101 s ago — their confusability will be 100/101. But after a further 1 s of retention has past, the confusability of these items will have increased just to 101/102—a very small increase. Although these confusabilities will not translate directly into recall probabilities (because confusability with other items and omission error probabilities will also be important) it is nevertheless intuitively clear why forgetting rate is likely to decrease over time in a temporal ratio model such as SIMPLE. Crucially, this occurs as a natural consequence of the scale-invariant (ratio like) properties of a model, and makes no reference to consolidation.

Finally, we found that there were situations in which the forgetting curve of the model was most accurately described by an exponential curve over the first 15 seconds or so of retention, and a power law thereafter. Figure 1 shows this behaviour of the model, with a different curve (of different functional form) fitted to the first 15 seconds of retention of a five-item list (exponential curve; $R^2 = .99$) and subsequent retention (power function: $R^2 = .999$). Because such behaviour can emerge in a model in which the fundamental mechanism for forgetting remains unchanged with timescale, we suggest that such forgetting data cannot be used to mandate a distinction between different memory systems operating over the short and the long term.

In summary: we have used a temporal distinctness model of memory to show that (a) slower forgetting of older memories can readily be explained without recourse to consolidation, and (b) available forgetting data do not require the assumption of two distinct memory stores with correspondingly different forgetting functions at different timescales.

4. Recency to primacy shift

We next consider instances in which memory performance actually *improves* over time. At first glance, such improvements are readily and naturally explained by consolidation, and they thus constitute a particular challenge for alternative models. We focus on one manifestation of performance-improvement over time known as the *recency to primacy shift.* When memory for a short list is tested immediately, the most recent items are almost always advantaged (unless their recall is postponed, as when memory for serial order is required). When a delay intervenes between presentation and test, the recency effect is reduced or abolished. Of greatest interest is the fact that performance on early list items may occasionally be *better* after a filled retention interval than on an immediate test (e.g. Bjork, 2001).

To clarify, the recency to primacy shift can refer to three different phenomena that are illustrated in Figure 2. Line A (identical in both panels) represents a typical pattern of performance on immediate testing – an extended recency gradient is seen. The remaining lines depict possible serial position curves after a filled delay. In all cases the recency effect is reduced. Panel 1 depicts two cases where performance on the primacy item(s) improves in absolute terms relative to their immediate performance: Line B shows the case where performance on the primacy item(s) has improved in absolute terms compared to immediate performance but not relative to later list items, and line C shows the case where primacy is increased in both absolute and relative terms after the delay. In Panel 2, line D shows the case where greater relative but not absolute primacy is observed after the delay, and line E shows the case in which performance on the primacy items does not improve after a delay in either absolute or relative terms.

All five patterns have been found in the data. The extent to which patterns C and D occur is controversial, but our main concern here is with any case (such as that seen in

B) in which performance on the primacy items improves over time. An initial demonstration of the recency-primacy shift (Wright, Santiago, Sands, Kendrick, & Cook, 1985) has been influential, and the absolute increase over time in performance on the primacy items—that is, patterns B and C in Figure 2— has been found with a number of studies, species, and methodologies (Wright, 2007). We note that in at least some cases, however, the phenomenon has not proved robust. The recency to primacy shift was found by Korsnes et al. (1996) and by Korsnes and Magnussen (1996) in a serial-order memory paradigm (participants were presented with single items and required to respond with the serial position of that item), but Kerr, Ward, and Avons (1998) found that the effect could be explained in terms of response bias. Early findings of the recency to primacy shift in recognition memory (Neath, 1993; Neath & Knoedler, 1994) also failed to be consistently replicated by Kerr, Avons, and Ward (1999). However Knoedler, Hellwig, and Neath (1999) replicated the increase in primacy with a filled delay under a number of conditions, and Bjork (2001) reviews evidence of a shift towards primacy in a range of literatures.

In summary, although the evidence is mixed, we adopt as a working hypothesis the possibility that there are instances in which performance on primacy items increases over time in absolute terms. In free recall, such effects could reflect recall-order phenomena – if early-presented items are recalled first, they will be advantaged over recency items through experiencing less output interference. Many of the relevant data have however come from serial memory and recognition tasks (see above). Such data are clearly problematic for the concept of time-based trace decay, for it is hard to see how memory could actually improve over time in such accounts. But could those data not be handled quite naturally by a consolidation view? Although attractive at first glance, we argue against this possibility because whenever a benefit is observed for primacy items, recent items would have an equal opportunity for consolidation during the increased retention internal. Thus, consolidation could only lead to an increase in either absolute or relative primacy under the (intuitively implausible) assumption that consolidation starts slowly, but then proceeds at an increasing rate after some time has passed (because unless there is an increasing rate of consolidation, the early items could never overtake the later ones).

By contrast, the recency to primacy shift again sits naturally within a temporal distinctiveness framework. Neath and Brown (2006) applied the SIMPLE model to the recency-primacy shift in recognition memory to show how the effect could be understood in terms of relative temporal distinctiveness; here we illustrate with a more recent temporally extended version of the model. Brown, Chater and Neath (2008) extended the ratio-rule temporal distinctiveness model of Brown, Neath, et al. (2007) to take account of the fact that items have a temporal extension—that is, they take up a contiguous slice of the temporal axis rather than a single point on it (the initial model made the simplifying assumption that the temporal locations of items could be treated as point sources, and this simplifying assumption is problematic when rehearsal data must be accommodated). The extended model preserves the assumption that memories are represented in terms of their positions along a logarithmically-compressed timeline receding into the past, but additionally represents the proportion of the memory timeline taken up by each item. Even if each item has the same actual duration, recent items will occupy more of the timeline than will more temporally distant items, because the latter occupy a more compressed region of the temporal memory dimension. It is assumed that the probability of recalling an item is determined partly by the proportion of the timeline that it occupies; such an assumption is consistent with a number of scale-invariant memory effects, such as the result of Maylor, Chater, and Brown (2001) showing that memories are retrieved at the same rate whether from the last week, month, or year.

Figure 3 shows the compressed timeline for immediate memory for a four-item list and delayed recall of the same list. If as just suggested memorability is determined by the *proportion* of the memory timeline occupied by items, it is readily apparent that that performance on the primacy items can increase after the filled retention interval, because the proportion of the timeline occupied by, for example, the first item increases in absolute terms (numbers within black squares denote proportion of total time occupied by an item). In terms of Figure 2, the pattern illustrated in panel 1, line B would be produced (see Bjork, 2001, for an alternative account).

In summary, we have argued that the recovery in memory of items over time (a) is problematic for trace decay models of memory, and (b) falls out naturally from a temporal distinctiveness framework without recourse to consolidation mechanisms. We now go on to argue that the same account can shed light on behavioural phenomena that have previously been taken as evidence for consolidation.

5. Consolidation

We have framed our discussion around the central question of whether cognitive models are deficient through not acknowledging a role for consolidation. Thus far, our critique of indirect evidence for consolidation—viz. the shape of the forgetting function, situations in which performance improves over time— revealed that the data can be equally (or better) accommodated by a distinctiveness model that does not involve consolidation. In this concluding section, we tackle head-on the behavioural data most widely cited in support of consolidation and, for each, offer an alternative theoretical interpretation in terms of temporal distinctiveness.

Temporal gradient of retroactive interference. A key finding is that the effect of retroactive interference is greater when it follows the target material in close temporal proximity. Wixted (2004b) beautifully reviews the relevant literature, much of which dates back almost a century, and we refer the reader to his summary (see also M. T. Dewar, et al., 2007). The basic paradigm is illustrated schematically in Figure 4. The time-line is represented as time moving from left to right; i.e., the present is represented at the right hand side of the figure. The three solid blocks at the left hand end of each panel represent material to be learned, the shaded block represents interfering activity (whether similar or dissimilar to the to-be-learned material — the distinction is not needed for the point being made here) and the bar marked "recall" represents the time of retrieval. Thus the retention interval for the to-be-learned material (the solid blocks) is constant in all three panels of the figure. However the time at which the interfering material occurs varies. In the top panel, the interfering material occurs immediately after the to-be-remembered material has been presented; in the middle panel the interfering material occurs midway between the to-be-leaned material and the time of recall, and in the lowest panel the interfering material occurs just before the time of retrieval. Empirically, as summarised by Wixted, the amount of material that is remembered is low when the retroactively interfering material follows on immediately after the to-be-remembered items (panel a), and higher when a long temporal gap intervenes between the learning and the interference (panel b). Performance may drop again when the interfering material immediately precedes retention (panel c), although this pattern is not always seen (e.g. M. Dewar, Garcia, Cowan, & Della Sala, in press, who find that amesic patients benefit monotonically from delay of interfering material). The material giving rise to this generalisation is dispersed and cannot be reviewed here; we instead take Wixted's summary as our starting point.

The temporal gradient of retroactive interference (RI) is the reducing effect of intervening interference as it becomes more temporally distant from the to-be-remembered material. This gradient sits very naturally with a consolidation account: the consolidation of the original material is assumed to be interrupted by the interfering material to a greater extent if the interfering material follows closely upon it (less time is available for consolidation of the learned items). However, on its own, consolidation is insufficient to explain the inverted-U shape of the temporal effects of RI: A second process is required to explain the impairment that is sometimes associated with interfering material occurring just before retrieval; this process is thought to be the high degree of competition provided by the interfering material at retrieval. Thus, the piece of behavioural evidence most widely cited in support of consolidation actually requires more than consolidation to explain. By contrast, the entire temporal pattern of interference is naturally, and arguably more parsimoniously, consistent with a temporal-distinctiveness approach to memory. We now sketch how such an account could work.

A key feature of distinctiveness that was implied by our discussion so far, but not made explicit, is that temporally crowded items will be less discriminable, and hence harder to retrieve. Crucially, such interference is local (Neath, Brown, McCormack, Chater, & Freeman, 2006) – only items that occupy nearby locations along the temporal continuum will interfere with each other. This principle is used in SIMPLE to explain a number of phenomena, such as proactive interference (Keppel & Underwood, 1962; Underwood, 1957) and the release from PI with the passage of time (Loess & Waugh, 1967). Indeed, temporal separation reduces proactive interference in AB-AD paradigms (Keppel, 1964; Underwood & Ekstrand, 1967; Underwood & Freund, 1968) and also over short time periods (Alin, 1968; Kincaid & Wickens, 1970; Peterson & Gentile, 1965). Moreover, there is ample evidence that temporally isolated items (those with longer temporal gaps surrounding them during presentation) are sometimes more easily remembered. Temporal isolation confers a recall advantage in free recall (G. D. A. Brown, Morin, & Lewandowsky, 2006); running memory span (Geiger & Lewandowsky, 2008), and memory for serial order when report order is unconstrained (Lewandowsky, Nimmo, & Brown, 2008). Forward serial recall presents a clear exception to this pattern however: temporally isolated items show little or no advantage in recall in such tasks (e.g. Lewandowsky, Brown, Wright, & Nimmo, 2006).[3]

Temporal distinctiveness models, therefore, predict exactly the pattern noted by Wixted (2004b) — a greater interfering effect of material that is temporally proximal to either study or test— but without any reference to consolidation and on the basis of a single process that is at the heart of distinctiveness. We note that the effect of interference-test proximity is not always observed (M. Dewar, et al., in press); the extent to which interfering material presented just before test will reduce recall may depend on details such as similarity between irrelevant and learned material, but this remains a topic for further research.

Moreover, the time-scale invariant properties of a temporal ratio model like SIMPLE enable it to predict time-based release from the threat of interference at a number of different timescales.

It turns out that the remaining sources of evidence for consolidation cited by Wixted (2004b) are subject to the same parsimonious re-interpretation within the distinctiveness framework.

Effects of sleep on memory. Sleep research has been of central importance in theorising about consolidation (Meeter & Murre, 2004). It has long been known that a list

---

[3] The fact that temporal isolation plays virtually no role in conventional forward serial recall has been taken to imply that people pay no attention to temporal information in those tasks, but use other dimensions such as position, instead (Lewandowsky, et al., 2006). This is entirely consistent with the observation of Lewandowsky et al. (2004) that there is no temporal forgetting in short-term serial recall. SIMPLE accommodates both results by postulating that items are represented along multiple dimensions, only one of which is temporal, and that people pay no attention to time in many short-term tasks.

of words is better remembered if it is followed by a retention interval during which the learner sleeps than if the learning is followed by the same retention interval filled not by sleep but by normal daily activity (Jenkins & Dallenbach, 1924). Sleep is assumed to protect memory from interference (e.g. Ellenbogen, Hulbert, Stickgold, Dinges, & Thompson-Schill, 2006), and it has been suggested that this reflects active consolidation processes that occur during sleep (e.g. Born, Rasch, & Gais, 2006). In support of a consolidation view, sleep's protective benefits are particularly pronounced if it occurs right after study. For example, Ekstrand (1972) showed that retention after a 24-hour retention period that included 8 hours of sleep was better if subjects slept right after study (81% recall) than if they slept right before test (66%). These and related findings are typically taken as evidence for consolidation — it is assumed that the process of consolidation continues during sleep, and that this is particularly beneficial early on during retention, whereas it is partially interrupted by the typical mental activities that otherwise fill the retention interval.

Again however the temporal distinctiveness model offers an alternative perspective without recourse to the concept of consolidation. The two conditions are illustrated in Figure 5, which follows the same labelling convention as Figure 4. Panel a shows the sleep condition; the retention interval that follows learning is unfilled by any new learning activity. Panel b shows the potentially interfering material assumed to follow learning when the retention interval is not sleep-filled, and Panels c and d illustrate the Ekstrand (1972) procedure described above. The superior memory performance in the sleep condition is predicted by temporal distinctiveness models for just the same reason as the temporal gradient of interference is predicted – the to-be-remembered material is rendered temporally isolated, and hence more retrievable, by the following gap during which little or no mental activity occurs. Furthermore, temporal distinctiveness models will predict reduced memory under conditions such as those shown in panel c, because the learned material is less temporally isolated as a result of the interfering material that immediately follows it. Whether interference will occur when the interfering material immediately precedes test, as illustrated in panel d, may depend on the similarity of the interfering material to the target material (and hence how strongly it competes for recall).

Retrograde amnesia. Perhaps the most widely-cited evidence for consolidation comes from the temporal gradient of memory loss associated with retrograde amnesia, which is often known as the Ribot gradient (see e.g. Meeter & Murre, 2004). The basic phenomenon, associated with damage to the medial temporal lobes, is extensively documented, and involves a loss of memory for past events in a temporally graded manner such that temporally more distant memories are relatively preserved, and more recent memories are lost to a greater extent. A. S. Brown (2002) reports a metareview of 247 outcomes from 61 articles, which leads him to conclude that the temporal gradient of memory is monotonic (the impairment gradually and continuously reduces as memories become increasingly temporally distant) and extremely long-lived (extending even up to half a century).[4]

How might this characteristic pattern of data be explained without reference to consolidation? Here we sketch an account according to which (a) access to the temporal dimension in memory is relatively more important for temporarily recent memories, and (b) access to the temporal dimension is lost in retrograde amnesia, leading to (c) selective loss of recent memories. We illustrate with a simple simulation.

According to a model like SIMPLE, items can be seen as occupying point locations in multidimensional memory space; the temporal dimension (on which we have focussed in the present chapter) is but one of many. Crucially for present purposes, SIMPLE includes the assumption that differential attentional weightings may be given to different dimensions in memory (G. D. A. Brown, Neath, et al., 2007; Lewandowsky, Nimmo, et al., 2008). Specifically, greater attentional weighting will be given to whichever dimension in memory space is most useful for the task in hand. For example, consider a case where memory items are located along just two dimensions: A temporal dimension that becomes compressed as items recede into the past, and a second "item" dimension that acts as a kind of shorthand representation for all the non-temporal

[4] The fact that consolidation seemingly extends over a time period that exceeds the duration of the average human life-span throughout much of human history has been leveled as a criticism against this interpretation (Nadel & Moscovitch, 1997). It is difficult to see how a consolidation process of that duration could have evolved.

dimensions along which an item would be represented. It would make sense for the memory retrieval system to pay relatively greater attention to the temporal dimension when retrieving relatively recent items (which will have quite distinctive locations along the time dimension), and to pay relatively less attention to the temporal dimension and correspondingly greater attention to the other dimension as stimuli recede into the past and the temporal dimension becomes less useful for distinguishing items. This can be seen as akin to the process of a shift from episodic to semantic memory (G. D. A. Brown & McCormack, 2006), and we noted earlier how attention being directed away from the temporal dimension might underlie the selective immunity of immediate serial recall to temporal isolation effects (Lewandowsky, et al., 2006) and temporal forgetting (Lewandowsky, et al., 2004).

A distinctiveness model augmented with an attentional mechanism offers a potential account of the temporal gradient associated with retrograde amnesia without recourse to the concept of consolidation. The top curve in Figure 6 shows the probability (in SIMPLE) of recalling an item as it recedes into the temporal past (the timescale is arbitrary) under the system just described whereby progressively less weighting is given to the temporal dimension (and more weighting to the other dimension) in memory for older items.[5] It is evident that there is a strong recency gradient, as would be expected, such that more recent memories are more likely to be retrieved. However, what would happen if information about items' locations along the temporal dimension becomes degraded or unavailable? The lower curve shows the temporal retroactive gradient that could result. Because recent memories rely more on availability of temporal information, they suffer more when that information becomes unavailable.[6]

---

[5] Specifically, it is assumed that the attentional weight given to the temporal dimension during memory retrieval reduces as a linear function of the temporal distance of the to-be-retrieved memory.

[6] This account assumes that the greater weighting given to the temporal dimension for recent items is relatively fixed, i.e., that it is not possible in the absence of access to temporal information to pay correspondingly greater attention to non-temporal dimensions for recent memories.

The simple toy model confirms that a temporal-distinctiveness approach may offer an account of temporally-graded amnesia without reference to consolidation. Of course, we make no claim to a complete account, and we focus purely on the behavioural data; there is a considerable body of neurobiological evidence consistent with consolidation mechanisms (e.g. Squire, Stark, & Clark, 2004) which remains to be examined, as well as other data that may implicate consolidation processes (see e.g. Born, et al., 2006). However we do not believe that the neurobiological evidence is necessarily inconsistent with the cognitive-level accounts provided here. In any case, it is perhaps not implausible that a neurobiological underpinning could be given for disruption of time-based retrieval. Many brain regions show signal differences as a function of the temporal distance of memories, such that recent stimuli show greater responses than older stimuli (Woodard, et al., 2007). Furthermore, memory for context may become gradually more independent of the hippocampus over time (e.g. Wiltgen & Silva, 2007). The temporal distinctiveness model that we have adopted for present purposes operates at the level of cognitive principle rather than neurobiological process. Recent items are assumed to be more memorable because of their greater temporal distinctiveness. But what gives rise to this greater distinctiveness at a mechanism level? A number of models share the idea that memory involves associating items to a temporal-contextual signal of some kind (G. D. A. Brown, et al., 2000; Burgess & Hitch, 1999, 2006; Lewandowsky & Farrell, 2008). One suggestion is, for example, that the signal is made up of a combination of high frequency and low frequency oscillators. This signal is assumed to change gradually over time, such that nearby states of the signal are more similar to each other than are more temporally separated states. Thus, if retrieval follows close upon learning, the context signal has had little time to change, and recent items benefit from the overlap between learning context and retrieval context. The benefit that items receive will depend on their recency, with the advantage progressively reducing as items recede further back in time – a recency gradient. Disruption of such a signal – if hippocampal damage were assumed to cause such damage – could lead to retrograde amnesia along the lines discussed above.

In summary, we have suggested that many of the behavioural data that have been taken as evidence for consolidation may be open to explanation in terms of other mechanisms. The mechanism based on distinctiveness that we put forward here clearly represents one candidate worthy of further exploration. However, there are other models that have been identified as promising candidates by on recent work on short-term forgetting; for example, the SOB model of Farrell and Lewandowsky and colleagues (Lewandowsky & Farrell, 2008) has been identified in a rigorous model comparison as being best able to handle data on short-term forgetting (Oberauer & Lewandowsky, 2008). It remains to be seen whether it could rival the account of the present phenomena provided by SIMPLE.

6. Conclusion

We began with the observation that many current cognitive models of memory accord no role to consolidation failure as a cause of forgetting, although they often make reference to trace decay. We have argued that recent progress in memory modelling, combined with a reassessment of the empirical evidence, undermines the case for trace decay as a cause of forgetting. We have also argued that further behavioural evidence is likely to be needed if cognitive modellers to be convinced to include consolidation mechanisms in their models that currently lack them.

References

Alin, L. H. (1968). Proactive inhibition as a function of the time interval between the learning of the two tasks and the number of prior lists. *Journal of Verbal Learning & Verbal Behavior, 7*(6), 1024-1029.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396-408.

Atkinson, R. C., & Shiffrin, R. M. (1971). Control of short-term memory. *Scientific American, 225*(2), 82-&.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*(1), 83-100.

Barrouillet, P., & Camos, V. (2009). Interference: unique source of forgetting in working memory? *Trends in Cognitive Sciences, 13*(4), 145-146.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 35*(2), 317-333.

Bhatarah, P., Ward, G., & Tan, L. (2008). Examining the relationship between free recall and immediate serial recall: The serial nature of recall and the effect of test expectancy. *Memory & Cognition, 36*(1), 20-34.

Bjork, R. A. (2001). Recency and recovery in human memory. In H. L. Roediger, III, J. S. Nairne, I. Neath & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 211-232). Washington, DC: American Psychological Association.

Born, J., Rasch, B., & Gais, S. (2006). Sleep to remember. *Neuroscientist, 12*(5), 410-424.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*(2), 201-233.

Brown, A. S. (2002). Consolidation theory and retrograde amnesia in humans. *Psychonomic Bulletin & Review, 9*(3), 403-425.

Brown, G. D. A., Chater, N., & Neath, I. (2008). Serial and free recall: Common effects and common mechanisms? A reply to Murdock (2008). *Psychological Review, 115*(3), 781-785.

Brown, G. D. A., Della Sala, S., Foster, J. K., & Vousden, J. I. (2007). Amnesia, rehearsal, and temporal distinctiveness models of recall. *Psychonomic Bulletin & Review, 14*(2), 256-260.

Brown, G. D. A., & McCormack, T. (2006). The role of time in human memory and binding: A review of the evidence. In H. D. Zimmer, A. Mecklinger & U. Lindenberger (Eds.), *Binding in human memory: A neurocognitive approach* (pp. 251-290). Oxford, England: Oxford University Press.

Brown, G. D. A., Morin, C., & Lewandowsky, S. (2006). Evidence for time-based models of free recall. *Psychonomic Bulletin & Review, 13*(4), 717-723.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*(3), 539-576.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*(1), 127-181.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551-581.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*(4), 627-652.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 45*(2), 177-192.

Chechile, R. A. (2006). Memory hazard functions: A vehicle for theory development and test. *Psychological Review, 113*(1), 31-56.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.

Cowan, N., Beschin, N., & Della Sala, S. (2004). Verbal recall in amnesiacs under conditions of diminished retroactive interference. *Brain, 127*, 825-834.

Cowan, N., Elliott, E. M., Saults, J. S., Nugent, L. D., Bomb, P., & Hismjatullina, A. (2006). Rethinking speed theories of cognitive development - Increasing the rate of recall without affecting accuracy. *Psychological Science, 17*(1), 67-73.

Crowder, R. G. (1989). Modularity and dissociations in memory systems. In J. L. Roediger, III & F. I. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving.* (pp. 271-294). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112*(1), 3-42.

Della Sala, S., Cowan, N., Beschin, N., & Perini, M. (2005). Just lying there, remembering: Improving recall of prose in amnesic patients with mild cognitive impairment by minimising interference. *Memory, 13*(3-4), 435-440.

Dewar, M., Garcia, Y. F., Cowan, N., & Della Sala, S. (in press). Delaying interference enhances memory consolidation in amnesic patients. *Neuropychology*.

Dewar, M. T., Cowan, N., & Della Sala, S. (2007). Forgetting due to retroactive interference: A fusion of Muller and Pilzecker's (1900) early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex, 43*(5), 616-634.

Ekstrand, B. R. (1972). To sleep, perchance to dream: About why we forget. In C. Duncan, P., L. Sechrest & A. W. Melton (Eds.), *Human memory: Festschrift for Benton J. Underwood* (pp. 59-82). East Norwalk, CT,: Appleton-Century-Crofts.

Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with theories of sleep and memory: Sleep, declarative memory, and associative interference. *Current Biology, 16*(13), 1290-1294.

Geiger, S. M., & Lewandowsky, S. (2008). Temporal isolation does not facilitate forward serial recall - or does it? *Memory & Cognition, 36*(5), 957-967.

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

Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36*(2), 73-137.

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

Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition, 35*(7), 1675-1684.

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *American Journal of Psychology, 35*, 605-612.

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

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning & Verbal Behavior, 1*(3), 153-161.

Kerr, J. R., Avons, S. E., & Ward, G. (1999). Effect of retention interval on serial position curves for item recognition of visual patterns and faces. *Journal of Experimental Psychology: Learning Memory and Cognition, 25*(6), 1475-1494.

Kerr, J. R., Ward, G., & Avons, S. E. (1998). Response bias in visual serial order memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 24*(5), 1316-1323.

Kincaid, J. P., & Wickens, D. D. (1970). Temporal gradient of release from proactive inhibition. *Journal of Experimental Psychology, 86*(2), 313-316.

Knoedler, A. J., Hellwig, K. A., & Neath, I. (1999). The shift from recency to primacy with increasing delay. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 474-487.

Korsnes, M. S., & Magnussen, S. (1996). Age comparisons of serial position effects in short-term memory. *Acta Psychologica, 94*(2), 133-143.

Korsnes, M. S., Magnussen, S., & Reinvang, I. (1996). Serial position effects in visual short-term memory for words and abstract spatial patterns. *Scandinavian Journal of Psychology, 37*(1), 62-73.

Lewandowsky, S., & Brown, G. D. A. (2005). Serial recall and presentation schedule: A micro-analysis of local distinctiveness. *Memory, 13*, 283-292.

Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language, 54*, 20-38.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review, 11*(5), 771-790.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *Psychology of Learning and Motivation: Advances in Research and Theory, Vol 49, 49*, 1-48.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language, 59*(2), 200-222.

Lewandowsky, S., Nimmo, L. M., & Brown, G. D. A. (2008). When temporal isolation benefits memory for serial order. *Journal of Memory and Language, 58*(2), 415-428.

Lewandowsky, S., & Oberauer, K. (2008). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin & Review, 15*(5), 875-888.

Lewandowsky, S., & Oberauer, K. (in press). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory & Cognition.*

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009a). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences, 13*, 120-126.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009b). Response to Barrouillet and Camos: Interference or decay in working memory? *Trends in Cognitive Sciences, 13*(4), 146-147.

Loess, H., & Waugh, N. C. (1967). Short-term memory and intertrial interval. *Journal of Verbal Learning & Verbal Behavior, 6*(4), 455-460.

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 53*(1), 1-22.

Lovatt, P., Avons, S. E., & Masterson, J. (2002). Output decay in immediate serial recall: Speech time revisited. *Journal of Memory and Language, 46*(1), 227-243.

Maylor, E. A., Chater, N., & Brown, G. D. A. (2001). Scale invariance in the retrieval of retrospective and prospective memories. *Psychonomic Bulletin & Review, 8*(1), 162-167.

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*(3), 419-457.

Meeter, M., & Murre, J. M. J. (2004). Consolidation of long-term memory: Evidence and alternatives. *Psychological Bulletin, 130*(6), 843-857.

Meeter, M., & Murre, J. M. J. (2005). TraceLink: A model of consolidation and amnesia. *Cognitive Neuropsychology, 22*(5), 559-587.

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.

Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 29*(6), 1353-1380.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7*(2), 217-227.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition, 21*(5), 689-698.

Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review, 10*(2), 430-434.

Neath, I., & Brown, G. D. A. (2006). Simple: Further applications of a local distinctiveness model of memory *Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 46, pp. 201-243).

Neath, I., Brown, G. D. A., McCormack, T., Chater, N., & Freeman, R. (2006). Distinctiveness models of memory and absolute identification: Evidence for local, not global, effects. *Quarterly Journal of Experimental Psychology, 59*(1), 121-135.

Neath, I., & Knoedler, A. J. (1994). Distinctiveness and serial position effects in recognition and sentence processing. *Journal of Memory and Language, 33*(6), 776-795.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110*(4), 611-646.

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115*(3), 544-576.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*(4), 761-781.

Peterson, L. R., & Gentile, A. (1965). Proactive interference as a function of time between tests. *Journal of Experimental Psychology, 70*(5), 473-478.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*(1), 129-156.

Portrat, S., Barrouillet, P., & Camos, V. (2008). Time-Related decay or interference-based forgetting in working memory? *Journal of Experimental Psychology: Learning Memory and Cognition, 34*(6), 1561-1564.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93-134.

Rubin, D. C., Hinton, S., & Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology: Learning Memory and Cognition, 25*(5), 1161-1176.

Rubin, D. C., & Wenzel, A. E. (1994). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*(4), 734-760.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115*(4), 893-912.

Shiffrin, R. M., & Steyvers, M. (1997). Model for recognition memory: REM - Retrieving Effectively from Memory. *Psychonomic Bulletin & Review, 4*(2), 145-166.

Sikstrom, S. (2002). Forgetting curves: implications for connectionist models. *Cognitive Psychology, 45*(1), 95-152.

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience, 27*, 279-306.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*(6), 1589-1625.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*(1), 49-60.

Underwood, B. J., & Ekstrand, B. R. (1967). Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology, 74*(4), 574-580.

Underwood, B. J., & Freund, J. S. (1968). Effect of temporal separation of two tasks on proactive inhibition. *Journal of Experimental Psychology, 78*(1), 50-54.

Vallar, G., & Baddeley, A. D. (1982). Short-term forgetting and the articulatory loop. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 34*, 53-60.

Wickens, T. D. (1999). Measuring the time course of retention. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model* (pp. 245-266). Mahwah, NJ.: Lawrence Erlbaum Associates Publishers.

Wiltgen, B. J., & Silva, A. J. (2007). Memory for context becomes less specific with time. *Learning & Memory, 14*(4), 313-317.

Wixted, J. T. (2004a). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review, 111*(4), 864-879.

Wixted, J. T. (2004b). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*, 235-269.

Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science, 14*(1), 6-9.

Wixted, J. T., & Ebbesen, E. B. (1990). On the form of forgetting. *Psychological Science, 2*(6), 409-415.

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition, 25*(5), 731-739.

Woodard, J. L., Seidenberg, M., Nielson, K. A., Miller, S. K., Franczak, M., Antuono, P., et al. (2007). Temporally graded activation of neocortical regions in response to memories of different ages. *Journal of Cognitive Neuroscience, 19*(7), 1113-1124.

Wright, A. A. (2007). An experimental analysis of memory processing. *Journal of the Experimental Analysis of Behavior, 88*(3), 405-433.

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Cook, R. G. (1985). Memory processing of serial lsists by pigeons, monkeys, and people. *Science, 229*(4710), 287-289.

Figure captions

Figure 1. A typical forgetting curve produced by the SIMPLE model (see G. D. A. Brown, Neath, et al., 2007 for details). The dashed line shows the best fitting exponential function to the first 15 s of retention; the unbroken line shows the best fitting power function to the retention function after 15 s.

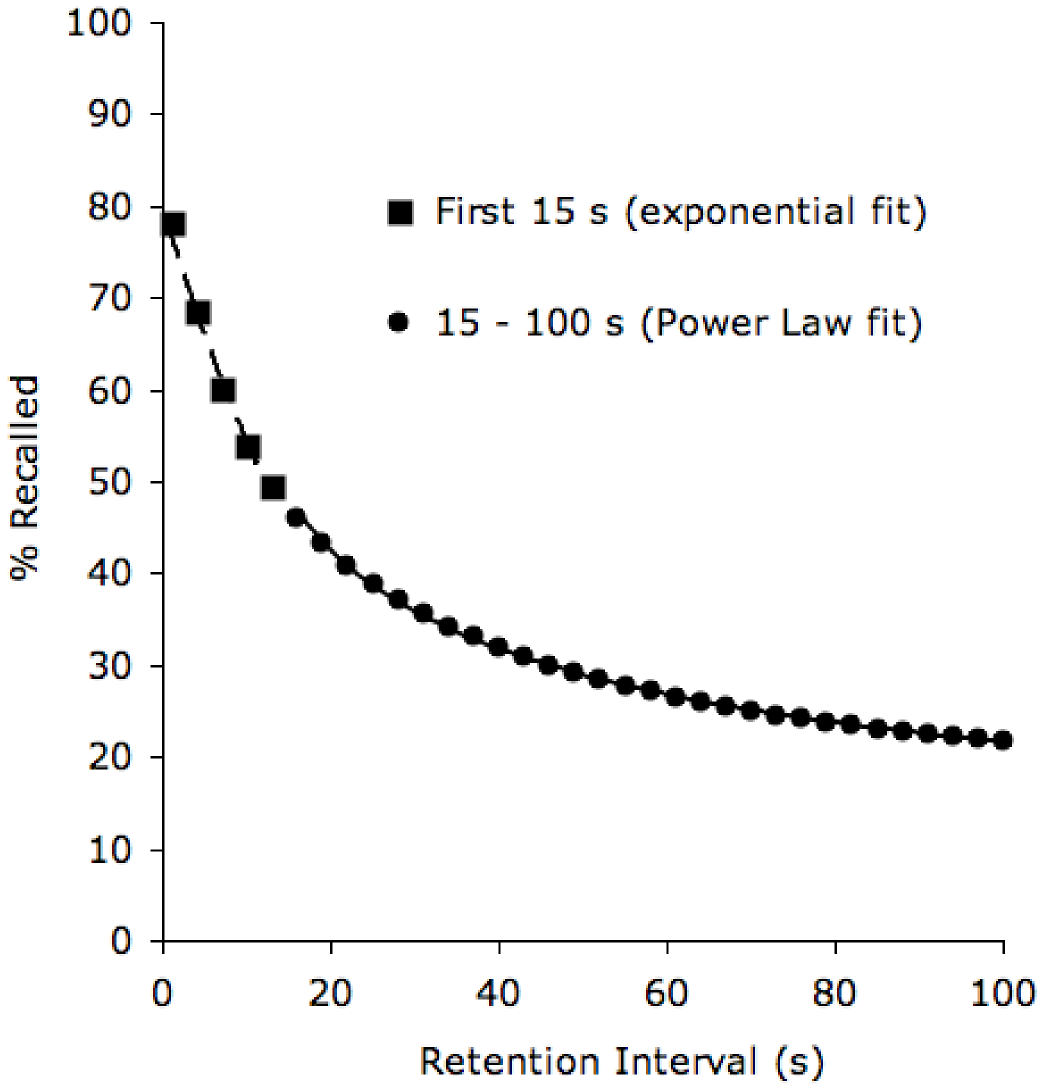
Figure 2. Illustration of various types of recency to primacy shift. Panel 1 illustrates increases in absolute performance on the primacy items; Panel 2 illustrates an increase in relative performance on the primacy items (see text for details).

Figure 3. Illustration of a compressed time-line in memory for immediate recall (top) and delayed recall (bottom). Each filled black rectangle represents an item, with the width of the item indicating the amount of the memory time-line that it occupies. Numbers within the squares represent the proportion of the timeline occupied by each item.

Figure 4. Illustration of the changing temporal distinctiveness of to-be-remembered items (filled black rectangles) as a function of the time of presentation of interfering material (filled shaded rectangles; panels a and b) and as a function of the time of sleep relative to the time of learning (panels c and d).

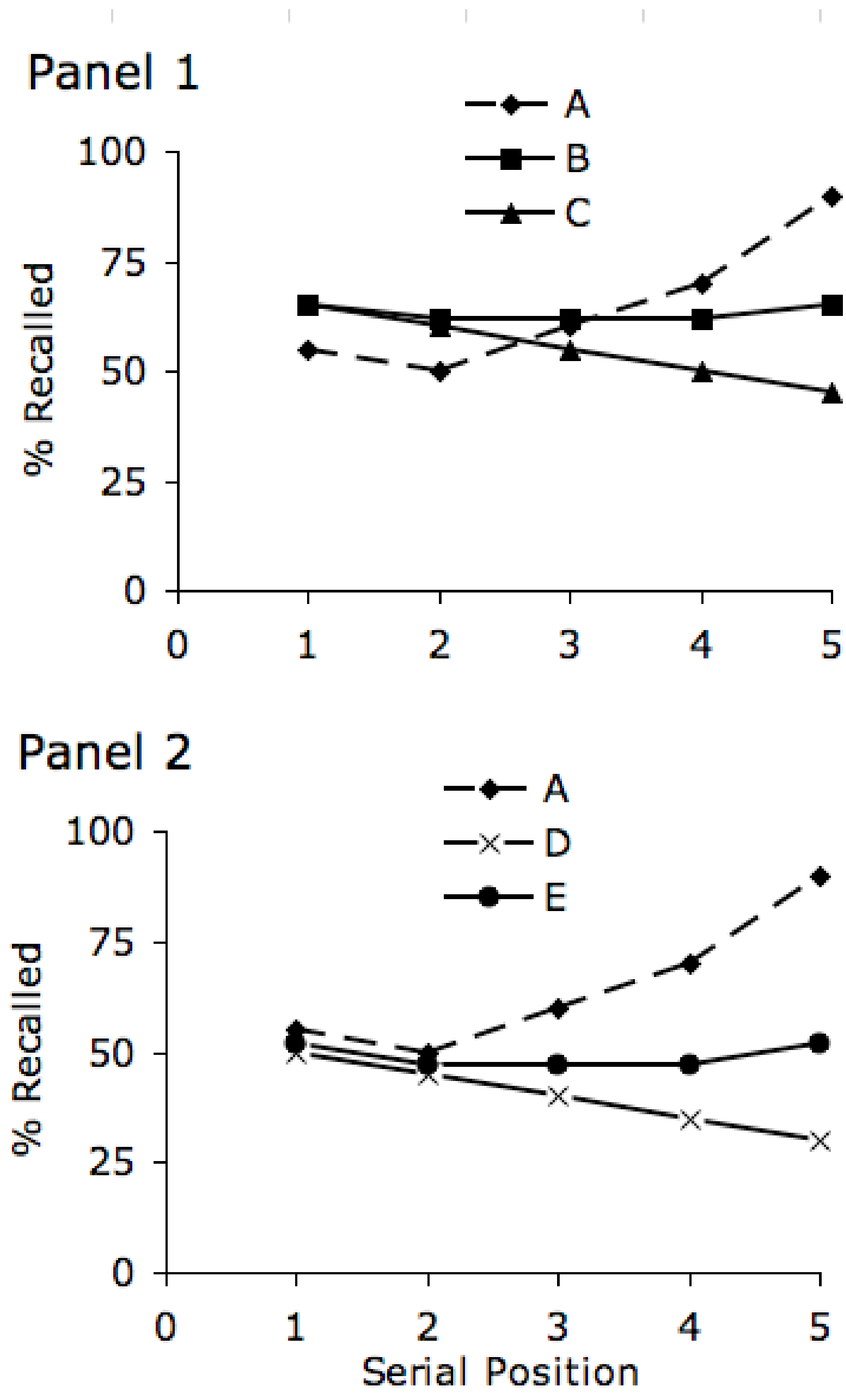
Figure 5. Illustration of how a typical recency gradient (top line) may be transformed into temporally graded amnesia (bottom line) if access to a temporal dimension in memory is lost (see text for details).

Figure 1



100
90
80
70
60
50
40
30
20
10
0
% Recalled
0
20
40
60
80
100
Retention Interval (s)
First 15 s (exponential fit)
15 - 100 s (Power Law fit)

Figure 2



Panel 1
A
B
C
100
75
50
25
0
% Recalled
0
1
2
3
4
5
Panel 2
A
D
E
100
75
50
25
0
% Recalled
0
1
2
3
4
5
Serial Position

Figure 3

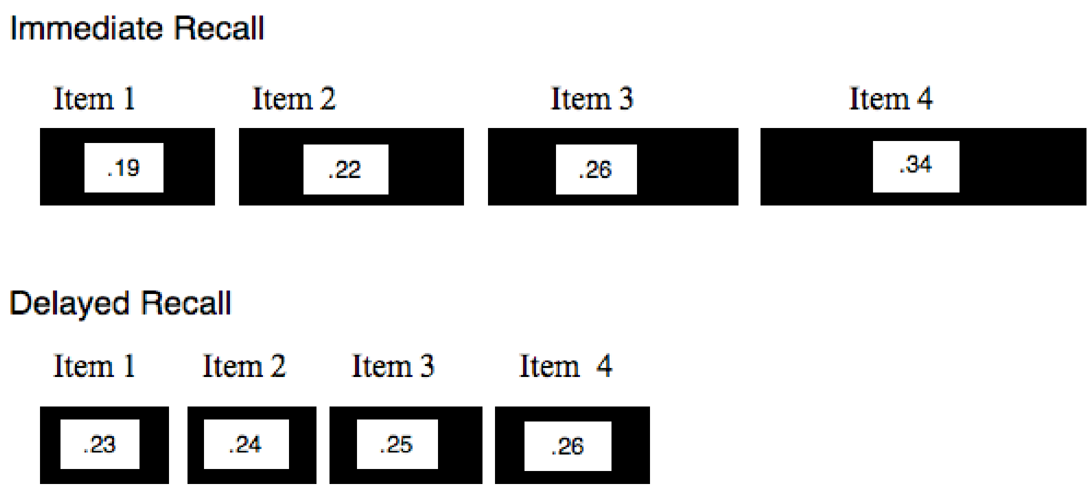

Immediate Recall
Item 1
Item 2
Item 3
Item 4
.19
.22
.26
.34
Delayed Recall
Item 1
Item 2
Item 3
Item 4
.23
.24
.25
.26

Figure 4

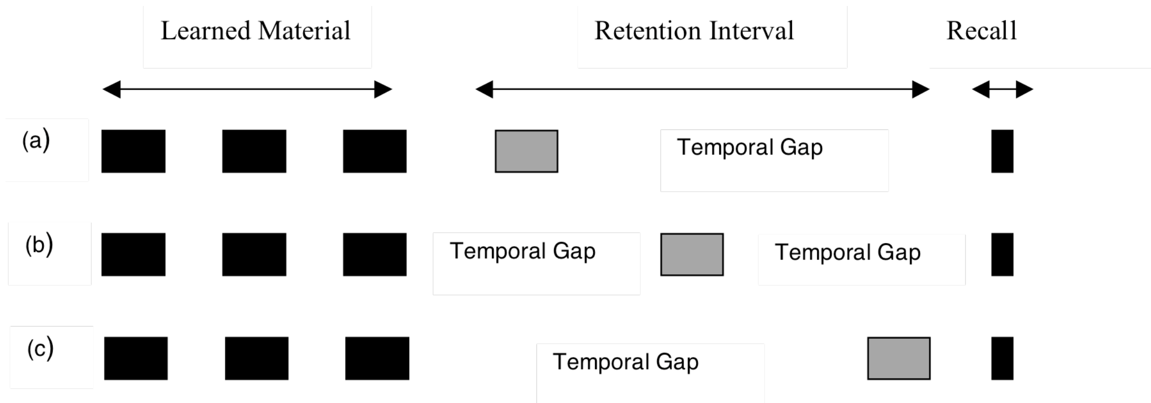

Learned Material
Retention Interval
Recall
(a)
Temporal Gap
(b)
Temporal Gap
Temporal Gap
(c)
Temporal Gap

Figure 5

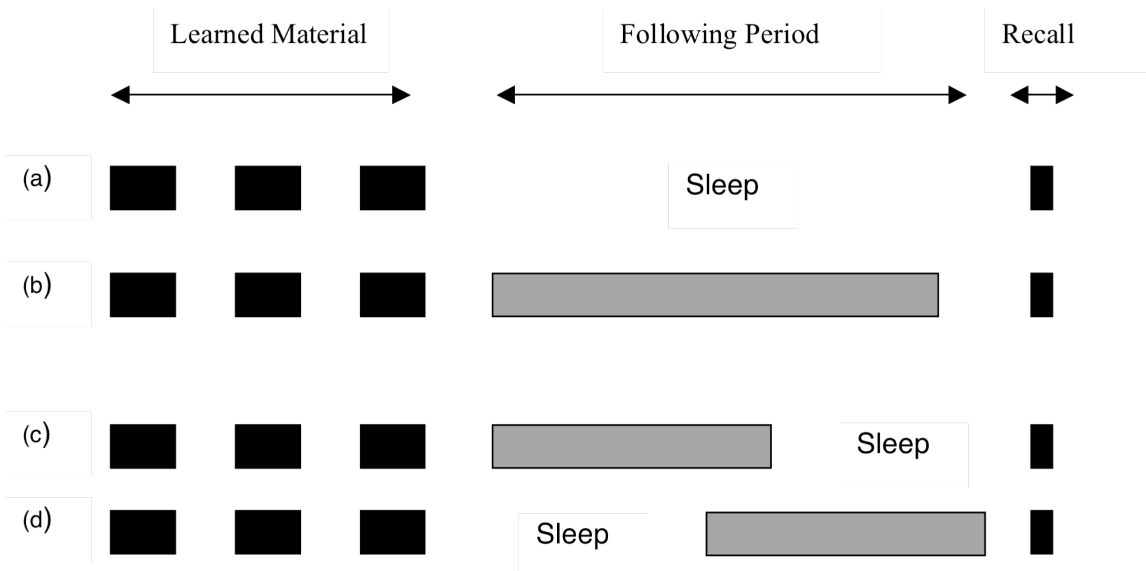
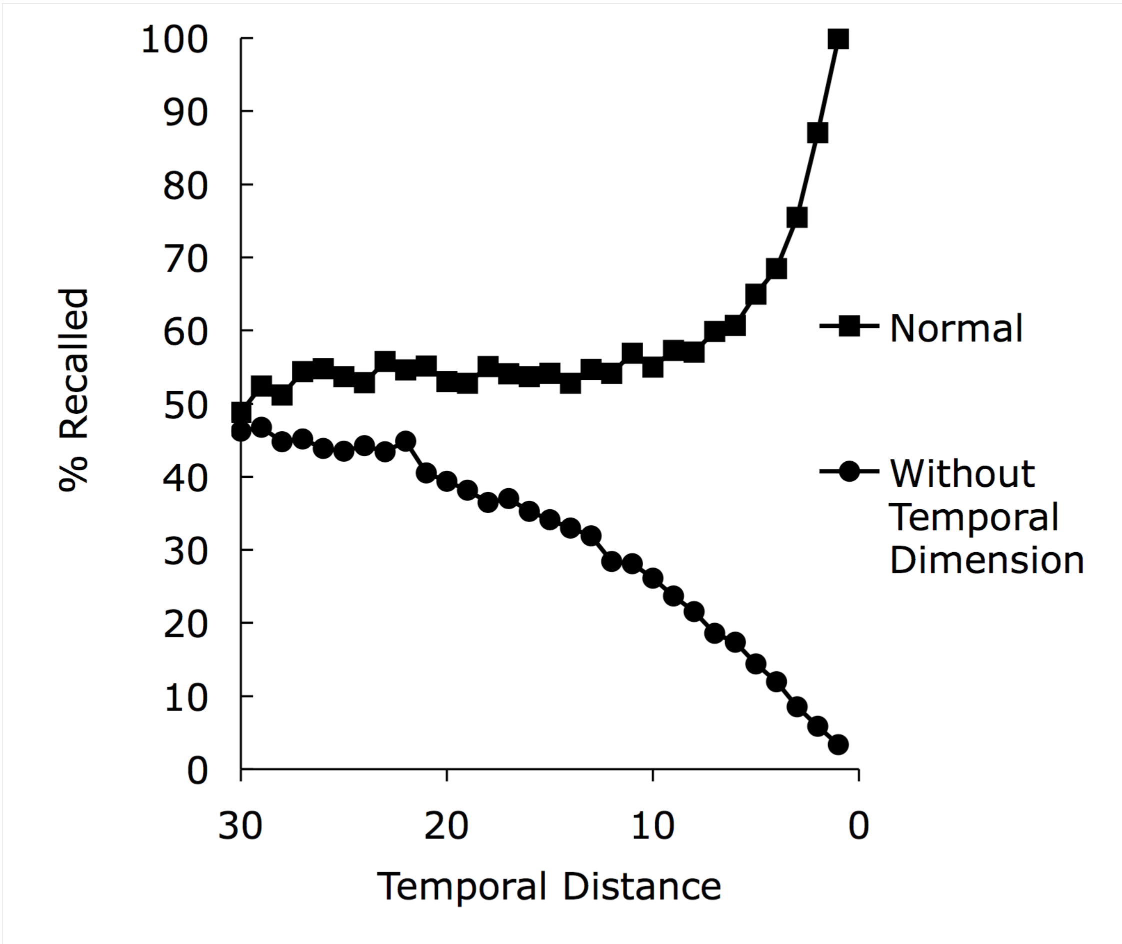

Learned Material
Following Period
Recall
(a)
Sleep
(b)
(c)
Sleep
(d)
Sleep

Figure 6


100
90
80
70
60
50
40
30
20
10
0
% Recalled
30
20
10
0
Temporal Distance
Normal
Without Temporal Dimension