A ratio model of scale invariant memory and identification

GORDON D. A. BROWN1, IAN NEATH2 and NICK CHATER1
1University of Warwick, Coventry, UK, and 2Purdue University, West Lafayette, Indiana

A model of memory and absolute identification is described. The model embodies 4 main claims: (a) Scale-independence: Similar mechanisms govern retrieval from memory over many different time scales; (b) Local distinctiveness: Performance on diverse absolute identification and episodic memory tasks is determined by interference from near psychological neighbors; (c) Temporal memory: Traces of items are represented in memory in terms of their temporal distance from the present; and (d) Interference-based forgetting: All memory loss is due to interference and not trace decay. The model is applied to many data on free recall, serial recall, and identification, and addresses new data on the identification of tones. The account emphasizes continuity in the retrieval principles involved in memory performance at all time scales, contrary to some models concerning distinctions between short-term and long-term memory.

It is characteristic of scientific laws to hold over a wide range of temporal, spatial, or physical scales. We would be surprised, for example, if Newton’s gravitational law, that the attractive force between two objects is inversely proportional to the squared distance between them, held for 1 mg objects but not 1 g or 1 kg objects, or for distances of centimeters but not distances of meters. Scientific principles are generally expected to hold at a wide range of scales (e.g. Barenblatt, 1996). Although laws may break down at extremes (e.g. for very small subatomic weights), the default search is for descriptions that are universal in that they apply over as wide a range of temporal and spatial scales as possible.

In developing models of human memory, however, it is widely assumed that different principles apply over different (short and long) time scales. Here we attempt to outline retrieval principles of human memory over both short and long time scales. We also relate time-invariant retrieval mechanisms to the mechanisms involved in identification of simple perceptual stimuli spread over either a narrow or broad psychophysical range. We term the resulting model SIMPLE (for Scale-Invariant Memory, Perception, and Learning). The model assumes that the confusability of items in memory is given partly by the ratios of the items’ temporal distances from the point of recall, and this emphasis on temporal ratios rather than absolute temporal durations
gives the model its scale-invariant properties.

We begin the paper with a brief review of serial position effects and scale-invariant effects in memory and absolute identification. The subsequent theoretical development of the model can be summarized as follows. Early attempts to provide a unified account of serial position effects in absolute identification and memory tasks (Murdock, 1960) had difficulty accounting for isolation effects in absolute identification (Bower, 1971). We show how a simplified exemplar theory (Nosofsky, 1986) can shed light on the problematic isolation effects in absolute identification while also capturing the scale-invariant serial position effects observed in the same tasks. The model, SIMPLE, is applied to a novel experimental analog of primacy and recency effects in absolute identification. This paves the way for a return to Murdock’s program of linking accounts of serial position effects in memory and identification paradigms. We then show that SIMPLE captures some scale-invariant and serial position effects in serial recall and free recall if it is assumed that (a) episodic memories are arrayed along a dimension representing temporal distance from the point of retrieval, (b) the retrievability of an item is inversely proportional to its summed confusability with other items in memory, and (c) the confusability of items along a temporal dimension is given by the ratio of the temporal distances of those items at the time of recall.

The model is used to account for many (although not all) data that have previously been assumed to reflect trace decay or the operation of a separate short-term memory, even though the model assumes that the same interference-based forgetting principles apply at all time scales. Thus the approach marks a return to the search for unitary principles of memory (Keppel & Underwood, 1962; Melton, 1963). The resulting model has affinities with a number of previous theoretical approaches, brings them together, resolves some of the difficulties, and argues that the combined approach can help to provide a unified perspective on a range of memory phenomena. The model can be viewed variously as: (a) an extension of Murdock’s (1960) distinctiveness theory that accommodates time-based and local neighborhood effects, (b) a generalization and extension of early temporal discriminability and ratio models of memory (e.g. Baddeley, 1976; Bjork & Whitten, 1974; Crowder, 1976; Glenberg & Swanson, 1986), (c) the addition of a temporal dimension to a simplified exemplar model (Nosofsky, 1986, 1992), (d) the extension into the domain of temporal memory of the Feature Model’s use of the Luce choice model for cue-driven recall (Nairne, 1990), (e) an extension of Neath’s (1993a, 1993b) temporal distinctiveness model to allow isolation effects and primacy effects to be accounted for, or (f) a more analytic and abstract version of recent oscillator-based and contextual overlap models of memory for temporal order (e.g. Brown, Preece, & Hulme, 2000). More generally, following Gallistel (1990), the model places time and temporal interference at the heart of memory and relates memory retrieval to perceptual discriminability.

**BASIC ISSUES**

**Serial Position Curves**

Serial position effects have been seen as evidence for distinctions between long-term and short-term memory, as evidence for the greater distinctiveness of end-series items, and as evidence for a relation between serial recall and absolute identification. They are therefore central to the concerns of the present paper. Bowed serial position curves, showing reduced memory for mid-series items, are obtained in free recall, serial learning, probed serial recall, location memory, and some recognition memory tasks (for reviews see Crowder, 1976; Lansdale, 1998, McGeoch & Irion, 1952; Murdock, 1974). Serial position
effects are also observed in absolute identification experiments in which items must be identified on the basis of their position along a single dimension such as frequency or amplitude (Attnave, 1950; Berliner, Durlach, & Braida, 1977; Braida & Durlach, 1972; Durlach & Braida, 1969; Kornbrot, 1978; Lacouture, 1997; Murdock, 1960; Woodworth & Thorndike, 1900). Serial position effects are also found in retrieval from long-term memory (Baddeley & Hitch, 1977; Bjork & Whitten, 1974; Glenberg, Bradley, Kraus, & Renzaglia, 1983; Healy, Havas, & Parker, 2000; Healy & Parker, 2001; Nairne, 1991; Pinto & Baddeley, 1991; Roediger & Crowder, 1976; Watkins, Neath, & Sechler, 1989) and in absolute identification experiments whether the stimuli to be identified span a narrow or wide psychophysical range (Lacouture, 1997; this paper). Many separate and incompatible theoretical accounts of these serial position effects are evident in the literature. Can a unified explanation be provided for these ubiquitous effects? Murdock (1960) answered affirmatively for the cases of identification and simple serial recall, suggesting that in both identification and recall items at the end of a series benefit through their greater relative distinctiveness. Accounts of serial position effects have subsequently diverged and proliferated. For example, in absolute identification, serial position effects have been attributed to response mapping (Lacouture & Marley, 1991, 1995), end anchor effects (Berliner & Durlach, 1973), response bias in combination with choice model application (Lansdale, 1998; Nosofsky, 1985), selective attention to end items (Luce, Green, & Weber, 1976), neighborhood density (Krumhansl, 1978), increased memory variance for mid-list items (Nosofsky, 1997), error-minimizing decision strategies (Balakrishnan, 1997), or dynamic criterion setting and response availability within a Thurstonian framework (Treisman, 1985; Treisman & Williams, 1984).

In memory, in contrast, early accounts of primacy and recency focused on proactive and retroactive interference respectively (e.g. Foucault, 1928; cited in Murdock, 2001). Primacy effects have been (non-exhaustively) attributed to additional rehearsal of early-list items (Rundus, 1971), more distributed or more recent rehearsal of early-list items (Modigliani & Hedges, 1987; Tan & Ward, 2000), various measures of distinctiveness (Johnson, 1991; Murdock, 1960), anchoring effects (Feigenbaum & Simon, 1962; Glanzer & Dolinsky, 1965), inhibitory processes (Hull, 1935; Lepley, 1934) or reducing encoding throughout a list (Brown et al., 2000; Farrell & Lewandowsky, 2002; Lewandowsky & Murdock, 1989; Page & Norris, 1998). Recency effects over short time-scales have been (non-exhaustively) attributed to primary memory (Waugh & Norman, 1965) or to response selection processes (Farrell & Lewandowsky, 2002; Lewandowsky, 1999; Lewandowsky & Farrell, 2000; Lewandowsky & Murdock, 1989), whereas long-term and short-term recency effects have been attributed to various temporal discriminability mechanisms (Baddeley, 1976; Baddeley & Hitch, 1993; Bjork & Whitten, 1974; Crowder, 1976; Glenberg et al., 1983; Glenberg & Swanson, 1986; Neath, 1993a,b; Tan & Ward, 2000). Thus in recent years the goal of developing a unified approach to serial position effects in memory and absolute identification has effectively been abandoned. Even within traditional serial and free recall paradigms, different explanations for primacy effects, short-term recency effects, and long-term recency effects survive in parallel.

Scale Invariant Effects in Memory and Perception

In addition to focusing on serial position effects in a range of tasks, the present paper aims to model data (from both absolute identification and memory paradigms) that hold over many different scales. What reason is there to believe that such an account is either
necessary or possible? We noted earlier that scale-invariant explanations are often preferred in science. Here we describe scale-invariant phenomena in memory and absolute identification, then emphasize the importance of scale-invariance for psychology more generally (Chater & Brown, 1999). The concept of scale-invariance is not always used consistently, and has technical usages that we do not require here, but in intuitive terms can be taken to characterize any system whose characteristics are similar independently of the scale of measurement (for accessible introductions see Bak, 1997; Flake, 1998; Mandelbrot, 1982; Schroeder, 1991). Scale-invariant physical structures are said to be statistically self-similar, in that small parts of an object are on average similar to the whole. The central point here is that the statistical structure of a data set (such as the degree of wiggliness of a river) provides no information as to the scale that is being examined. Natural-world examples are provided by structures such as clouds, or the coast of Norway - in both cases an observer would be unable (in the absence of any other cues) to distinguish between a small nearby structure and a larger distant structure. In human memory, an analogous example detailed below is the ratio-dependent nature of recency effects such that examination of a recency effect will not in itself be informative regarding the scale of the effect (i.e., whether it is a long-term or a short-term recency effect).

Scale Invariance in Memory

Evidence for scale-independence in human memory may take a number of different forms. Particularly relevant data come from results that implicate similar processes operating over different timescales; such results motivate models like the one presented here in which relative amounts of time, not absolute amounts of time, influence retrievability.

First of all, evidence for power-law forgetting would be consistent with scale-independence in memory performance over time. If the forgetting curve does follow a power-law, then the probability of recall will depend on \( T^a \) where \( T \) is the amount of time since an episode was learned and \( a \) is a constant. To the extent that the time course of human memory loss does more or less closely follow a power-law, as has been suggested by a number of researchers (see e.g. Anderson & Schooler, 1991; Rubin & Wenzel, 1996; Wicksted and Ebbesen; 1991, 1997), forgetting may be seen as scale-independent. Many researchers have however claimed that forgetting curves are not best described by a power law (for recent examples see Rubin, Hinton, & Wenzel, 1999; Wickens, 1999; cf. also Myung, Kim, & Pitt, 2000); we address this issue below.

Further evidence consistent with scale-independence has come from the study of recency effects. Although recency effects disappear after a filled retention interval (Glanzer & Cunitz, 1966; Postman & Phillips, 1965), the effect reappears if the spacing between presented items is increased (Bjork & Whitten, 1974) and is seen when retrieval from LTM is required (Baddeley & Hitch, 1977; Pinto & Baddeley, 1991; Sehulster, 1989). More generally, the size of the recency effect appears to depend on the log of the ratio between the IPI (inter-presentation interval between the items) and the RI (retention interval) (Glenberg et al. 1980; Nairne, Neath, Serra, & Byun, 1997), at least to a remarkably close degree (see also Baddeley, 1976; Bjork & Whitten, 1974). The relation holds even when the IPI to RI ratio varies over many orders of magnitude, from milliseconds to

\[ y \propto x^a \]

Power-law relationships are evidence for scale-invariance in a way that other functions (such as logarithmic or exponential ones) are not, because a power-law relationship holds independently of the measurement scale. Thus if \( y \) is proportional to \( x^a \), the same relationship will hold even if all \( x \) values are multiplied by a constant (although the constant of proportionality will change, \( a \) will not).
weeks. Empirically, the ratio rule means that it would be impossible for an observer to judge from the size of a recency effect whether that recency effect arose from recall of a list of items presented 10 s apart and followed by a 50 s retention interval, or a list of items presented 1 s apart and followed by a 5 s retention interval.

A third line of evidence for scale-independence in memory is the fact that the proportion of errors produced for each serial position during serial learning remains constant even in the face of considerable variations in degree of learning, IPI, time between trials, familiarity or meaningfulness of the material to be remembered, or individual differences in the learners (Braun & Heymann, 1958; McCrary & Hunter, 1953); this is the Hunter-McCrary Law. Here the shape of the error distribution in serial learning provides the observer with no evidence about the absolute level of performance. We present representative data below.

Fourth, serial position effects in rather different tasks, such as order reconstruction, remain qualitatively (and sometimes quantitatively) unchanged at different time scales (e.g. for the dimensions of position-within-list and list-within-trial: Nairne, 1991; data below). Similar qualitative scale-invariant features are evident in the data from grouping experiments, where items at the beginning and end of each group are better recalled, echoing the primacy and recency effects for the list as a whole (Frankish, 1985, 1989; Ryan, 1969a, 1969b; Hitch, Burgess, Towse, & Culpin, 1996).

Fifth, the pattern of transposition errors in order reconstruction tasks remains remarkably constant across timescales varying over many orders of magnitude, from milliseconds to weeks (see Huttenlocher, Hedges, & Prohaska, 1992; Nairne, 1991, 1992; Neath, 1998; see Brown et al., 2000, for a summary). A sixth line of evidence comes from the study of memory for relative recency. Underwood (1977) asked participants to recall the dates of events that occurred between four months and seven and half years in the past, and found that the greater the time separating two events, the less likely those events were to be recalled in the wrong relative order. Hacker (1980) found similar effects for items separated by just a few seconds. Seventhly, there is evidence for invariance over the time-frame of recall. Maylor, Chater, and Brown (2001) asked participants to recall events from the past day, week, or year. The cumulative response probabilities were indistinguishable across the three conditions. Finally, scale-invariance is also evident in recall-order effects in free recall (Howard & Kahana, 1999).

In summary, there is considerable evidence that many important properties of memory are scale-invariant, in that similar effects are evident at many different timescales (see also Melton, 1963; Naimre, 1992; 1996). An emphasis on scale-invariance resonates with recent models of scale-invariant processes in animal learning; Gallistel and Gibbon (2000) describe a model of scale-invariant conditioning that is closely related to the model of human memory proposed here. Models of timing also emphasize scale-invariance (e.g. Killeen & Taylor, 2000; Wearden, 1994).

**Scale Invariance in Absolute Identification**

Scale-invariance is also observed in absolute identification. In a typical absolute identification task, participants are exposed to a set of stimuli arrayed along some dimension such as pitch, weight, area, or amplitude (e.g. nine tones of different frequencies, in the experiments we report below). A label is associated with each stimulus. The labels may be numbers (e.g. 1 through 9), with the number for each item corresponding to the item’s ordinal position on the continuum, or may be arbitrary (e.g. the names of different colors may be associated with the different stimuli, in which case the task essentially becomes more
akin to paired associate learning). Participants are then exposed to individual stimuli in random order and required to identify them with the correct label. Feedback regarding the correct response is normally given after each trial.²

We present data from several absolute identification studies below; here the central (and counterintuitive) point is that identification performance for pairwise-discriminable items is often almost unaffected when the spacing of items along the perceptual scale is increased by a constant factor (see, e.g., Alluisi & Sidorsky, 1958; Eriksen & Hake, 1955; Garner, 1962; Miller, 1956; Pollack, 1952; Shiffrin & Nosofsky, 1994). This provides a clear example of scale-invariance. Further scale-invariance is evident in the serial position effects obtained in absolute identification experiments, for such curves have essentially the same form no matter what the range of the relevant perceptual dimension that is used (Neath, Brown, McCormack, Chater, & Freeman, 2006).

Scale Invariance and Psychological Explanation

Scale-invariant phenomena appear to call for scale-invariant models. Is a bias towards scale-invariant models inconsistent with memory data? In the study of human memory, it has been assumed that the data do not permit scale-invariant explanations. Distinctions between the memory retrieval principles that operate over short and long time scales have a long and distinguished history (Atkinson & Shiffrin, 1968; James, 1890/1950; Waugh & Norman, 1965) and are still widely accepted (see, e.g., Baddeley, Gathercole, & Papagno, 1998; Cowan, Wood, & Borne, 1994; Gathercole, 1999; Izawa, 1999). In particular, time-based trace decay is assumed to be an important factor underpinning forgetting over short timescales, with interference being assumed to predominate over longer timescales. Our exploration of scale-invariant retrieval principles is not of course intended to rule out the assumption of a separate STM (although we do suggest that much existing evidence can be interpreted within a scale-invariant framework). In focusing on aspects of memory that appear similar over widely varying time scales we attempt a new perspective on an old tradition of suggesting that the same interference-based principles apply over both short and long time scales. We call this the scale-invariant memory assumption. The present proposals regarding scale-invariance in human memory and absolute identification particularize a more general claim that scale-invariance in cognitive and perceptual function may reflect scale-invariance in the structure of the environment (Chater & Brown, 1999; cf. Anderson, 1990; Shepard, 1987a).

Time, Identification, and Categorization

Our concern is with temporal scale invariance in the case of memory, but with invariance over stimulus range in the case of absolute identification. A common perspective is sought - do the same principles govern the discriminability (and hence retrievability) of items in memory as govern the discriminability of stimuli from one another in absolute identification paradigms? More specifically, do serial position effects arise for the same reasons in memory and in identification (Murdock, 1960)? The relationship between absolute identification and serial and free recall remains unclear, and in recent years different theoretical and modeling approaches

² Absolute identification is typically characterized by a near-absence of performance improvement over large amounts of learning (e.g. Shiffrin & Nosofsky, 1994) and this, as well as the procedure involving use of unidimensional stimuli and an ordered response continuum, distinguishes it from paired-associate learning. Absolute identification and paired-associate learning are distinguished from absolute judgment in which sensory magnitude judgments are given without feedback, although an associative component may be involved in absolute judgment (Haubensak, 1992; Wedell, 1996; but see Parducci, 1992).
have been taken in explaining these different paradigms. Current models of absolute identification, categorization, and recognition performance account for a range of empirical data to a high level of precision (e.g. Ashby, 1992; Ashby & Perrin, 1988; Erickson & Kruschke, 1998; Estes, 1994; Kruschke, 1992; Kruschke & Johansen, 1999; Lamberts, 1995; Nosofsky, 1986; Nosofsky & Palmeri, 1997). In many respects such models seem more advanced than current models of serial and free recall. However the insights embodied in models of identification and classification have not generally been applied to traditional serial and free recall memory paradigms. Here we suggest that this is partly because multidimensional scaling models of categorization have not included time as an important dimension underpinning memory retrieval. Although models can allow for the differential availability in memory of exemplars, the relation between temporal factors and memory/exemplar availability has not been widely explored. However in recent models of memory, and of serial recall in particular, time has been accorded a central role (e.g. Altmann & John, 1999; Anderson & Matessa, 1997; Anderson, Bothell, Lebiere, & Matessa, 1998; Brown et al., 2000; Burgess & Hitch, 1996; 1999; Houghton, 1990; Neath, 1993a, 1993b; see also Hintzman & Block, 1971; Hintzman, Block, & Summers, 1973) and in more general terms it has long been argued that time is an important dimension underpinning memory organization and retrieval (e.g. Gallistel, 1990; see Brown & Chater, 2001, for a recent review; Friedman, 2001, for an alternative perspective). Here we apply simple principles of the type previously explored mainly in categorization and identification models to serial recall and free recall paradigms via the addition of a temporal distance dimension into such models.

SUMMARY OF THE MODEL

The SIMPLE model of identification and memory retrieval is intended to apply to all cases where items must be identified on the basis of their position along a psychological dimension. In an absolute identification experiment, for example, participants must retrieve identifying labels for each item in a set of stimuli that differ in terms of their position along a single dimension (such as size, brightness, or frequency). In memory we assume that the dimension is often temporal and specifies the time elapsed since the to-be-remembered item was learned.

The basic idea is that items are more distinctive, and hence both more memorable and easier to identify, to the extent that they are located in sparsely-populated regions of psychological space. The effect can be illustrated by the (imaginary) task of identifying individuals on the basis of their height alone. Figure 1a shows seven individuals, labeled A through G, who differ only in terms of their height. After being shown all the individuals and their identifying labels, participants are shown one individual and are required to reproduce the letter-label for that individual (see “Test Phase” in Figure 1a).

The model assumes that stimulus magnitudes (here, people’s heights) are represented in memory as logarithmically transformed values (Figure 1b). When a single person is subsequently shown for identification, the person’s height acts as the cue for retrieval of the person’s identifying label. The retrieval cue (the perceived height of the test person) is compared with the remembered heights of all the people in the set. The similarity of the perceived test height to each remembered height is assumed to be negatively related to the distance between test height and memory height on the internal psychological scale.

For example, in accommodation to such issues as the dangers of averaging data over participants: Ashby, Maddox, and Lee (1994); Maddox, (1999).

For example, via the M parameter in the Nosofsky and Palmeri (1997) EBRW model.
(i.e., heights that differ by a small amount will be seen as very similar). The probability of identifying a given test person as a particular remembered individual is determined by the similarity of the test height to that particular person’s remembered height relative to the similarity of the test height to all remembered heights.

Intuitively, this task will be harder if the test individual has a similar height to other individuals, i.e., if it has many near neighbors in height space. For example, D will be harder to identify correctly at test than A or B, because D is similar in height to C and E. D’s similarity to A, B and G will have less influence on task difficulty. Items near the
ends of the series will have an advantage as they have fewer near neighbors. It is this intuition — that near neighbors will be most influential in determining task difficulty — that is captured by the “local distinctiveness” model presented in the present paper. Identification will be difficult to the extent that many possible responses are associated with similar dimensional values to a test item; memory retrieval will be difficult to the extent that a retrieval cue points to a locally crowded region of psychological space.

The above assumptions are essentially similar to those embodied in exemplar models of categorization and identification, but have not generally been applied to human memory for serial order and free recall. Here, the assumption regarding memory is that in traditional free recall and serial recall paradigms the specificity of a retrieval cue for an item depends on that item’s time of occurrence relative to the time of retrieval and that, analogously to the absolute identification case, memories that are locally distinctive along the temporal dimension will be better recalled. This is seen as analogous to the task of picking out an individual from others that are differentiated only in terms of their position along a line, in terms of a remembered position (Crowder, 1976). Again, items in crowded regions of the array will be difficult to pick out on the basis of their position. The model we develop goes beyond previous distinctiveness and ratio models of memory (e.g. Baddeley, 1976; Bjork & Whitten, 1974; Crowder, 1976) in that it can account for grouping effects and primacy effects as well as recency effects and in that it predicts memory interference from several memory items, not just nearest neighbors. For memory the model can be expressed straightforwardly in terms of temporal ratios. The confusability of any two items along the temporal dimension is just the ratio of their temporal distances raised to some power. For example, the confusability of items that occurred 4 s and 5 s ago would be \((4/5)^c\) where \(c\) is a free parameter. The confusability of items that occurred 1 s and 2 s ago would be smaller, being \((1/2)^c\). The probability of recalling a given item will be inversely proportional to its summed confusability with all other items. More specifically, for simple paradigms involving memory for lists of unrelated items the model predicts that the probability \(R_i\) of recalling an item \(i\) from a set of \(k\) other items if the temporal distance of the item is \(T_i\) is given by:

\[
P(R_i | T_i) = \frac{1}{\sum_{j=1}^{k} \text{Ratio}(T_i, T_j)^c}
\]

where \(\text{Ratio}(x, y)\) is the smaller of \(x\) and \(y\) over the larger. A similar expression predicts absolute identification of stimuli that vary along a single dimension such as frequency or weight.

Additional assumptions are needed to allow for omissions, hierarchical effects, and the importance of non-temporal dimensions that specify items’ positions in phonological or semantic space. However it is the emphasis on ratios of temporal durations in determining forgetting, rather than on absolute amounts of time, that gives the model its scale-invariant properties and distinguishes it from trace decay models as such models are traditionally conceived.

**IMPLEMENTATION**

We now describe the implementation of SIMPLE in detail, and show how the simple description in terms of ratios, as given above, derives from more basic underlying assumptions and can be extended. The initial model makes the Fechnerian assumption
that $M_x \propto \log(S_x)$, where $M_x$ is the value that item $x$ will have on an internal scale if it is veridically perceived and $S_x$ is the value of item $x$ on the relevant physical dimension (which may be amplitude, temporal duration, temporal distance, frequency, or any other dimension that is psychophysically well-behaved within the range used).  

In absolute identification, is assumed that after the initial presentations of the items with their identifying labels (items are presented several times in a typical experiment, prior to the test phase in which items are presented without labels for identification by participants), accurate representations of each item’s value on the sensory scale are stored in memory. To adopt the person-identification analogy, suppose that the participant encounters three people of different heights (A = 164 cm; B = 181 cm; C = 200 cm). After several presentations of heights associated with labels, these would be accurately represented in memory as logarithmically transformed $M_i$ values: 5.10 for A; 5.20 for B, and 5.30 for C. It is assumed that these mnemonic representations do not decay or degrade over time – there is no trace decay.

On presentation of a test person for identification (e.g., A), the model assumes that participants have an accurate (in the mean) internal representation of the test person’s dimensional value (5.10), in addition to the veridical representations of the originally presented psychological scale values (5.10; 5.20; 5.30). The task is then one of determining which of the originally seen people is the best “hypothesis” to is the best “hypothesis” to explain the “datum” of the test person’s height. We therefore make the standard assumption, following a version of the similarity choice model (Luce, 1963; Shepard, 1957), that the probability of responding $R_j$ given a stimulus $S_i$ will be determined by the similarity of the memory representation of $S_i$ ($M_i$) to the memory representation of item $j$ ($M_j$) relative to the similarity of the internal representation of $S_i$ to all other values stored in memory. That is:

$$P(R_j | S_i) = \frac{(\eta_{i,j})^\gamma}{\sum_{k=1}^{n}(\eta_{i,k})^\gamma}$$ (2)

where $n$ is the number of items in the set and $\eta_{i,j}$ is the similarity between $M_i$ and $M_j$ in memory. In other words, the probability that a given test item will be correctly identified will depend on the similarity of that item’s internal representation at test to that item’s memory representation relative to the test item representation’s similarity to all other representations in memory. (Here the items are assumed to be equiprobable and response bias is ignored.) The parameter $\gamma$ governs how deterministic responses are – high values of $\gamma$ imply that participants will respond very consistently in judging a given input; values of $\gamma$ near zero imply that participants will make highly variable responses with only a slight preference for the ‘correct response’ (Ashby & Maddox, 1993). In all the simulations in this paper, we set $\gamma$ to 1.0, and the parameter can be ignored for present purposes. The above model suffices for a basic illustrative account of serial position accounts in absolute identification and for most paradigms in-

---

5 Natural logarithms are used throughout this paper.

6 Thus described the model makes the “direct mapping” assumption (Laming, 1997) of a direct link between the absolute physical magnitude of a stimulus and its resulting value on an internal psychological scale; we show below how this assumption can be relaxed in order to accommodate the fact that accurate absolute rather than merely relative perception of absolute magnitudes is not possible when items are presented in decontextualised isolation.

7 In the simple version of the model, the response determinism parameter will have the same effect as $c$ and so there is no advantage in including the two parameters separately.
Involving memory for serial order; additional threshold mechanisms are required to allow for the possibility of omission errors in free recall and we introduce these as necessary.

Crucially, and in line with many other models, it is assumed that similarity \( \eta_{i,j} \) falls off as a decreasing function of the separation between any two representations \( M_i \) and \( M_j \) on the internal scale (Shepard, 1987b). Thus for the absolute identification case where stimuli differ along just one dimension:

\[
\eta_{i,j} = e^{-c|M_i - M_j|^q}
\]  

(3)

where \( c \) is a constant and \( q \) is 1.0 for an exponential function relating similarity to distance and 2.0 for a Gaussian function relating similarity to distance. Use of this function, which is widely used to relate similarity to distance in psychological space for separable stimuli (e.g. Nosofsky, 1986; Shepard, 1957, 1987b) has the effect that items that are very close on a psychological scale have a similarity approaching 1.0 (because \( \eta_{i,j} = 1 \) when \( |M_i - M_j| \) \( q \) = 0 ), whereas items that have more psychologically distant representations from one another have a similarity that approaches zero as the psychological distance becomes greater (because \( \eta_{i,j} \) tends to 0 when the psychological distance \( |M_i - M_j| \) tends to infinity). The rate at which similarity reduces with psychological distance is given by the parameter \( c \). This function relating similarity to psychological distance becomes more strongly related with each other as \( c \) increases (because \( \eta_{i,j} \) tends to 0 when the psychological distance \( |M_i - M_j| \) \( q \) tends to infinity). The rate at which similarity reduces with psychological distance is given by the parameter \( c \). This function relating similarity to psychological distance is shown in Figure 2, for four different values of \( c \). The effect of using the exponential similarity-distance function is that the probability of correctly identifying a given item is determined most strongly by the similarity of its psychological scale value to those of its immediate neighbors, as suggested by the person-identification example. More distant neighbors have less influence, with their influence reducing as a function of their psychological distance from the target.

In applications to episodic memory and absolute identification below, \( q \) was set to 1.0 (i.e. the function relating similarity to psychological distance was assumed to be exponential). This constrains and simplifies the model considerably, and in combination with the earlier assumption that internal psychological magnitudes are logarithmically transformed stimulus energy values, i.e. \( M_x \propto \log(S_x) \), allows the model to be expressed very simply in terms of ratios. This proceeds as follows.

Rewriting (3) with \( q \) and \( q \) set to 1.0, we express the similarity between two memory representations as:

\[
\eta_{i,j} = e^{-c[M_i - M_j]}
\]

(4)

and using \( M_x = \log(S_x) \), this becomes:

\[
\eta_{i,j} = e^{-c[\log(S_i) - \log(S_j)]}
\]

(5)

and therefore:

\[
\eta_{i,j} = e^{-c[\log(S_i/S_j)]}
\]

(6)

We use the fact that \( \log(S_i/S_j) \) =
\[
\log(S_i/S_j) \text{ if } S_i \geq S_j, \text{ and } |\log(S_i/S_j)| = \log(S_j/S_i)
\]
if \( S_i < S_j \), to obtain

\[
\eta_{i,j} = \left( \frac{S_j}{S_i} \right)^c \tag{7}
\]

if \( S_i \geq S_j \) and

\[
\eta_{i,j} = \left( \frac{S_i}{S_j} \right)^c \tag{8}
\]

if \( S_i < S_j \).

The interpretation of this is straightforward. Given values of \( S_i \) and \( S_j \), say 5 and 15, the similarity between them is the smaller value divided by the larger (here \( 5/15 = 1/3 \)) raised to the power \( c \). If the values are separated by a large ratio (e.g., 1 and 10) the similarity will be small (0.1\(^c\)); conversely if they are identical, their similarity will be maximal, at 1, equal to their ratio. Let us define a function \( \text{Ratio}(x, y) \), which divides the smaller of \( x \) and \( y \) by the larger. Then we can write:

\[
\eta_{i,j} = \text{Ratio}(S_i, S_j)^c \tag{9}
\]

In other words, the similarity of two memory values is some power of their ratio. Substituting into Equation 2, we obtain:

\[
P(R_j | S_i) = \frac{\text{Ratio}(S_j, S_i)^c}{\sum_{k=1}^{n} \text{Ratio}(S_i, S_k)^c} \tag{10}
\]

Thus response probabilities can be expressed purely in terms of the ratios (raised to the power \( c \)) of stimulus energy values. When memory is being modeled, and the \( S_i \) values are temporal distances from the point of retrieval, the resulting emphasis on ratios of temporal distances brings out the close relationship of SIMPLE to the temporal discrimination models of Baddeley (1976) and Bjork & Whitten (1974) in addition to other ratio models of memory. Neath and Brown (in press) contains a detailed analysis of the relation between different distinctiveness models.

Of the possible free parameters in the simple model outlined above, just one, the slope parameter \( c \) (which specifies the steepness of the function relating psychological distance to similarity) was varied in the initial simulations below. The account therefore differs from previous application of exemplar-based models to absolute identification data in the absence of a multidimensional scaling stage to establish the structure of the underlying psychological space and in the absence of response bias parameters (see Nosofsky, 1992, for a review). These simplifications, although substantial, permit insight into many relevant serial position data while at the same time allowing ready generalization to the cases of serial and free recall.

We first apply the model to absolute identification data, with a focus on serial position effects as a function of the distribution of stimuli within a range. This application provides the foundation for subsequent application of the same model to serial position effects in memory.

### SERIAL POSITION EFFECTS IN ABSOLUTE IDENTIFICATION

**Data from Murdock (1960)**

Although similar to previous exemplar-based models of absolute identification, the SIMPLE model differs from previous distinctiveness models of identification. For example, Murdock’s influential (1960) model assumes that the distinctiveness of items is determined by their psychological distances from all items in the set of to-be-discriminated items (“global distinctiveness”), while SIMPLE proposes in contrast that distinctiveness is given primarily by an
item’s psychological distances only from other items that are nearby in psychological space ("local distinctiveness") (see Neath & Brown, in press). Can SIMPLE nonetheless shed light on the relative serial position effects accounted for by Murdock’s global distinctiveness model?

Figure 3. Relative proportion correct identification observed in four absolute identification experiments (data adapted from Murdock, 1960); solid line shows fit of the SIMPLE model (see text for details).

Representative results from several experiments are illustrated in Figure 3, where the data comes from identification of amplitudes (panels a-c; data from Murdock, 1960); and weights (panels d and e, data from Murdock, 1960). Similar serial position functions are observed when the underlying dimension is frequency (new data we report below), line length (Bower, 1971), area (Eriksen & Hake, 1957), position along a semantic continuum (DeSoto & Bosley, 1962; Pollio & Deitchman, 1964, cited in Bower, 1971), spatial position (Ebenholtz, 1963; Jensen, 1962), brightness (Bower, 1971), or serial/temporal position in a memory list (many studies, reviewed below). In other studies from our own laboratory we have found similar serial position effects in the absolute identification of temporal durations, wooden rods varying in length, and numerosities (Neath et al., 2006).

The results of applying SIMPLE to the serial position effect data shown in Figure 3 are shown on the same figure. The single free parameter, \( c \), was set at 0.386 (weights) and 0.075, 0.027, and 0.047 (loudnesses 1 through 3 respectively).\(^8\) The fits of the

\(^8\) Values of \( c \) were calculated on the basis of natural log transformations of raw stimulus energy values
model to the data were generally good. In terms of fit SIMPLE does as well as, or better than, the global distinctiveness model on these data, although comparisons are not meaningful as the local model has one free parameter whereas the global model has none. Fit statistics are therefore not reported.

The fit of the model to the data on the identification of weights is systematically deviant. Performance is not good enough on the larger weights relative to performance on the smaller weights. Inspection of Figure 3d suggests that the deviation between model and data may be due to the use of an inappropriate zero for the scale. A weight of zero is assigned a magnitude of zero, but this fails to take into account the weight of the hand and arm (see e.g. Laming, 1997). A better fit is obtained if a constant value is added onto all weights. The results of adding a constant of 5.7 (lb.) are shown in Figure 3e; \( c \) was set to 2.62.

Why does the model predict superior performance for items near the ends of the relevant continuum? Whenever an item is presented at test for identification, its internal representation will on the average be most similar to the memory representation of the reference stimulus of the same magnitude, but it will also be partially similar to the memory representations of stimuli nearby in magnitude. These near neighbors will effectively compete to control the response, and the more similar neighbors that there are, the lower will be the probability of correct identification of the test item. When items at or near the end of the perceptual continuum are presented for identification, there will on average be fewer similar competitors, and the probability of correct identification will be higher. This explanation in terms of “edge effects” has much in common with explanations proposed by Estes (e.g., 1972), Houghton (e.g., 1990), Treisman (1985) and others; we postpone detailed consideration.

In summary, SIMPLE does reasonably well in accounting for the basic serial position curves illustrated in Figure 3. The detail of the fits must be treated with caution as no information on response bias tendencies in the data (e.g. the sometimes-observed tendency for mid-series responses to be preferred in absolute identification) was available; any response biases in the model’s output were removed in a parameter-free way as explained below. The observed serial position effects do not depend on this however.

**Isolation Effects in Absolute Identification**

The next step is to show that SIMPLE can account for data that are problematic for earlier distinctiveness models. These concern isolation effects. As Bower (1971) and Murdock (1974) both noted, a feature of global distinctiveness models such as that of Murdock (1960) is that they may never predict that a mid-list item is more discriminable than its immediate neighbors, no matter how distinctive the item might, intuitively, seem to be. However Bower observed that under certain conditions it must be possible to isolate a mid-series item from its neighbors in psychological space to the extent that it would become more discriminable than its neighbors. For example, in a series of tones of frequencies 400, 420, 440, 540, 640, 660, and 680 Hz., the middle tone (540 Hz.) is widely separated from its immediate neighbors, and intuition suggests that this isolated item will be more easily identified than will its immediate neighbors. We have recently provided experimental confirmation of Bower’s (1971) intuition:

---

9 We note that this figure is somewhat larger than those found to maximize conformity to Weber’s Law when weight discrimination is the task (see Laming, 1997).
mid-series items can indeed be identified more accurately than their immediate neighbors if they are sufficiently isolated in psychological space (Neath et al., 2006). The results of these experiments (and appropriate control conditions) are reproduced in Figure 4, together with the fit of the SIMPLE model described below. The only free parameter was $c$; this was set to 6.4 (frequencies); 3.6 (rod lengths); 5.8 (weights); 2.9 (numerocities), and 3.7 (line lengths).

The emphasis on local rather than global neighborhoods means that the isolated items have few near neighbors and hence are identified relatively accurately. In terms of the person identification example, the analogy is with the difficulty of naming a person who is very similar in height to two other people, compared with the ease of identifying the same person when there is no immediate height-neighbor (see Figure 1). The heights of the tallest and shortest people in the set will have much less effect on identification than will the heights of the near neighbors. Thus a simple confusability-based account, while incomplete, captures an important aspect of the serial position curve data that form the current focus.

Scale Invariance in Absolute Identification

As a further step before extending the model to memory phenomena, we apply the model to a set of absolute identification data collected to examine spacing effects and scale-independence effects on serial position curves in absolute identification (for details, see Neath & Brown 2005; for further discussion of fitting SIMPLE to these data, see Neath & Brown, 2006). The spacing effects are central to the argument because they will be used to show that the asymmetrical serial position effects typically observed in memory (e.g. greater recency than primacy) can also emerge in absolute identification.

Experiment 1, therefore, examined participants’ ability to identify individual tones from sets of nine tones varying in frequency. Three sets of nine tones were used; the difference between them was the spacing between them - the spanned range varied from 140 Hz (narrowly-spaced set of nine tones) to 290 Hz (widely-spaced set). In each of the three sets, stimuli were separated by equal ratios throughout the range. The stimulus spacings for Experiments 1 (narrow, medium, and wide) and 2 (primacy and recency) are illustrated in Figure 5.
The results of Experiment 1 are shown in Figure 6 (top three left panels). Clear symmetrical serial position effects were observed, along with characteristic positional uncertainty gradients indicating that even if a stimulus was not identified correctly the response that was given tended to be close to the correct one. Consistent with scale-invariance, and with previous research, overall accuracy did not vary significantly across the three conditions (although, as occurs in other studies, there was a trend towards better performance on more widely separated items). There was little learning over the course of the experiment, consistent with the interpretation of the experiment as absolute identification rather than paired associate learning: averaged over all five conditions in Experiments 1 and 2, performance was 44% correct during the first half of the trials, and 46.5% during the second half.

Can the SIMPLE model accurately fit these results, and be extended to account for the full response confusion matrix (data presented in the left hand column of Figure 6 below)? This raises three related issues. We deal with each of these in turn here as similar issues are relevant to several subsequent simulations of serial memory and free recall tasks.

Some earlier models, based on “relative distinctiveness”, give predictions only for the relative performance that will be seen for identification of different items in a series. But why do absolute levels of performance typically remain approximately invariant over changes in stimulus range? Accounts have focused on increased stimulus noise (Durlach & Braida, 1969) or increased memory noise (Gravetter & Lockhead, 1973) as explanations for the reduction in local discriminability that accompanies extension of the range spanned by stimuli to be identified. Nosofsky (1983a) argued that both stimulus noise and memory noise in-
crease with the range covered by the to-be-identified stimuli. Here we simply assume that the $c$ parameter, which can be seen as incorporating both types of variation, increases linearly with stimulus range. The rationale is as follows. One way of conceptualizing the scale-independence of absolute identification is in terms of the “stretchiness” of the internal scale used to represent the items in question. It is as if participants have a fixed quantity of dimensional capacity, which can be stretched or contracted to accommodate the task demands most efficiently. In the case of the “narrow” tones in the present experiment, for example, the scale will cover a range of 563-420 = 143 Hz, whereas in the “wide” condition the range is 652-363 = 280 Hz. The resolution of the scale in absolute frequency space will depend upon the task demands. This flexibility has the advantage that sensitivity can be task-dependent, and that adaptation can occur (cf. Helson, 1964; Parducci, 1995).

The concomitant disadvantage is that participants have no direct access to information about absolute magnitudes. Instead only relative judgments can be made - so whenever isolated stimuli are presented, and a judgment about the magnitude of the resulting sensation must be made, there is always an implicit comparison baseline of some kind (see Helson, 1964). This baseline may be a stimulus presented on a previous trial (Jesteadt, Luce, & Green, 1977), or may be some complex amalgam of remembered experience (Stewart, Brown, & Chater, 2002).

We refer the reader to Laming’s comprehensive summary of evidence for this and related general claims; several researchers have examined the specific effects of prior stimuli in absolute identification tasks (e.g. Luce, Nosofsky, Green, & Smith, 1982; Mori & Ward, 1995; Ward & Lockhead, 1970). From this point on we simply assume the inherent relativity of judgments of absolute magnitudes. Thus in modeling the data of Experiment 1, we assumed that the frequency of each test tone is judged in relation to the frequency of the previously-presented tone (random presentation order means that each tone is equally likely to be preceded by each other tone over the duration of the experiment). Participants were always provided with feedback; we assume that the participant’s response on trial $n$ would be based on the perceived difference between tone 2 and tone 5. When perception of this difference is not subject to noise or distortion, we can continue to talk as if the absolute frequencies of the tones are available to participants. A more detailed model, in which the difference-perception process is assumed to be subject to noise and systematic distortion, is developed by Stewart et al. (2002) and shown to account for sequential effects such as assimilation and contrast (e.g. Ward & Lockhead, 1970, 1971). Here we use the simple one parameter model to facilitate understanding of the serial position effects.

*Relative and Absolute Magnitudes.* Laming (1997) has summarized many decades of psychophysical research, the results of which are consistent with the idea that participants are often unable to make reliable decontextualised judgments of absolute sensory magnitudes. Instead only relative judgments can be made - so whenever isolated stimuli are presented, and a judgment about the magnitude of the resulting sensation must be made, there is always an implicit comparison baseline of some kind (see Helson, 1964). This baseline may be a stimulus presented on a previous trial (Jesteadt, Luce, & Green, 1977), or may be some complex amalgam of remembered experience (Stewart, Brown, & Chater, 2002). We refer the reader to Laming’s comprehensive summary of evidence for this and related general claims; several researchers have examined the specific effects of prior stimuli in absolute identification tasks (e.g. Luce, Nosofsky, Green, & Smith, 1982; Mori & Ward, 1995; Ward & Lockhead, 1970). From this point on we simply assume the inherent relativity of judgments of absolute magnitudes. Thus in modeling the data of Experiment 1, we assumed that the frequency of each test tone is judged in relation to the frequency of the previously-presented tone (random presentation order means that each tone is equally likely to be preceded by each other tone over the duration of the experiment). Participants were always provided with feedback; we assume that the participant’s response on trial $n$ would be based on the perceived difference between tone 2 and tone 5. When perception of this difference is not subject to noise or distortion, we can continue to talk as if the absolute frequencies of the tones are available to participants. A more detailed model, in which the difference-perception process is assumed to be subject to noise and systematic distortion, is developed by Stewart et al. (2002) and shown to account for sequential effects such as assimilation and contrast (e.g. Ward & Lockhead, 1970, 1971). Here we use the simple one parameter model to facilitate understanding of the serial position effects.
**Response Bias.** Response bias can be fitted, explained as part of a model, or removed. Exemplar models of absolute identification often make use of a separate bias parameter for each item (e.g., Nosofsky, 1985). This approach increases the number of free parameters but allows precise control over response bias at the level of individual participants and individual serial positions. However, when a separate response bias parameter is estimated for each serial position, and/or when a transformed psychological scale derived via multi-dimensional scaling is used to provide input to the model, then empirically observed serial position effects could be attributed in the model to response bias in decision making, or to greater psychological spacing of stimuli at the extremes of the range, or to reduced response availability. In line with our general aim of emphasizing explanatory transparency at the possible cost of failing to capture every nuance of the data, we factored out response biases. Each response probability \( P(R_i|S_j) \) is divided through by the experimentally derived summed response probability \( R_i \Sigma \) with the result that all response probabilities sum to 1.0. The response probability matrix is then re-normalized so that the stimulus presentation probabilities return to unity, and the process is repeated iteratively until both stimulus probabilities and response probabilities are 1.0. Although this procedure can lead to distortion when major response biases are present, it causes no noticeable or theoretically significant distortion when biases are minimal and unsystematic and the same procedure is followed for both model and data. There was no systematic response bias in the data reported here. Responses 2, 6, and 7 were slightly but significantly preferred over responses 4, 5, and 9; there was an effect of spacing only on bias for response 3. This pattern contrasts with the occasional systematic finding that end responses are produced less often in absolute identification experiments and also in probed serial recall experiments (Underwood, 1977).11

This method of removing response bias (when such bias is small and unsystematic) permits an arguably more transparent account of serial position effects, in that symmetrical effects are unambiguously attributable to edge effects (response availability) and asymmetrical effects can unambiguously be attributed to the nature of the underlying dimensional representation. A disadvantage of the approach is that SIMPLE, like other models, offers no explanatory account of response bias effects in absolute identification (see Stewart et al., 2002, for the beginnings of such an account).

**Applying The Model To The Data.** The behavior of the model (Equation 10) is shown in the right-hand column of Figure 6 (the top three panels), where it is evident that a reasonable fit to the serial position curves and confusion matrices \( R^2 = 0.98 \) was obtained.12 There was one free parameter, \( c \), for the application of the model to the 243 data points. The value of \( c \) was set to 23.03 for the narrow condition, and scaled up in proportion to the highest:lowest frequency ratio for the other two conditions. We show below that a further 162 data points can be captured quite well without changing the parameter value. Thus the fit seems reasonable for such a

---

10 For example when participants are highly conservative for particular responses.

11 The model sometimes shows a small tendency to produce fewer extreme responses and more middle responses, as seen in some absolute identification and probed serial recall experiments, although the model’s performance is modulated in complex ways by overall level of performance.

12 Similar patterns are obtained whether residual summed squared error, or log-likelihood (calculated from estimated frequencies), or \( G^2 \) is minimized.
able for such a constrained model. In particular, the characteristic scale-invariant serial position effects, showing the advantage for items at the ends of the series, are observed. As before, the serial position effects arise because items near the end of the series are more “locally distinctive” - they have fewer near neighbors that can be similar to a presented test item. (The same account will be used below to address serial position effects in serial and free recall.) The model therefore illustrates an account of serial position effects in absolute identification that does not require any assumption of end-anchoring or differential attention to end-series items. The latter point deserves note, for it has been widely assumed that differential sensitivity for end-series items is needed to explain bow effects in absolute identification, and that an account of the type proposed here cannot suffice (e.g. Lacouture, 1997; Nosofsky, 1983b). For example, Treisman (1985) concludes that there are two quite separate “bow effects” – one concerning edge effects on proportion correct, and the other on differential sensitivity. The success of SIMPLE in accounting for the shape of the serial position curve in itself suggests that differential underlying sensitivity is not a significant factor in accounting for serial position curves. However, the main motivation for the assumption of a serial position curve in sensitivity has in any case come from the observation of serial position effects in $d'$ measures (e.g. Berliner et al., 1977; Durlach & Braida, 1969; Lacouture, 1997; Luce et al., 1982; Nosofsky, 1983b). But these data are ambiguous of interpretation; Stewart et al. (2002) have noted that bow effects in the $d'$ measure as it is typically calculated (Luce et al., 1982; Nosofsky, 1983b; see also Lacouture & Marley, 1995) can emerge without differential sensitivity in or attention to the endpoints of an underlying scale.

**Primacy and Recency Effects in Absolute Identification**

The first experiment demonstrated that the local distinctiveness model can address serial position effects in the identification of tones that are equally spaced (on a ratio scale), and of the fact that performance level and serial position curves are roughly scale-invariant. The main aim of Experiment 2 is to vary the spacing of items in an absolute identification task in order to test the prediction that asymmetrical serial position curves will thereby be obtained. To anticipate: the observation of asymmetrical serial position curves will be used to underpin subsequent accounts of primacy and recency effects.

Experiment 2 is also designed to rule out two alternative accounts of the serial position effects obtained in Experiment 1. First, a natural explanation of scale-invariant serial position curves in absolute identification is that the relevant distinctiveness-determining dimension is ordinal position in the series. If tones are encoded ordinal as “highest”, “second-highest”, and so on, the absolute values of the frequencies will make no difference to performance (assuming the tones are far enough apart to be discriminable: Bower, 1971). A parallel issue arises in considering serial recall from memory, where items may be ordered in terms of their location along either a positional or a temporal dimension. Experiment 2 was designed to rule out this ordinal-code interpretation of Experiment 1. Two sets of tones were created such that, according to the SIMPLE model but not an ordinal-code model, one set should lead to more accurate identification of the lowest numbered items (the ‘primacy’ condition) whereas the second should lead to more accurate identification of the

---

13 Calculation of log likelihood statistics reveal that statistically better fits can be achieved if additional parameters are incorporated, but we focus here on the simplest possible model and the main qualitative effects.
highest numbered items (the ‘recency’ condition). This was done by separating ordinal position and frequency differences. This can be achieved experimentally by increasing the separation on the frequency scale of tones at one or other end of the series, while holding the overall range constant. The frequency values of the tones that were used are illustrated in Figure 5.

Second, serial position effects as well as scale independence can be partially predicted by a simple account that we will call the “ordinal guessing” model. Assume that participants are able to judge only whether a given tone is “higher than” or “lower than” the tone presented on the preceding trial, and that participants then guess randomly between available responses. For example, participants hear tone 6 on trial n-1, and receive feedback (“that was tone 6”). Participants then hear tone 8 on trial n, and correctly judge this to be higher in frequency than the previous tone. The tone must therefore be tone 7, 8, or 9, and participants choose randomly from these responses, resulting in 33% correct responses. This process of accurate ordinal judgment followed by guessing will lead to serial position curves with advantages for the end-series items, because when end-series items are presented there will be fewer guessable responses and so the correct response will be picked more often by chance alone. (This assumes that each tone is preceded equally often by every other tone). As it is not intuitively obvious how this ordinal guessing process would translate into serial position curves, we implemented the model.

The results are shown in Figure 7a, where it can be seen that an excellent fit to the results of Experiment 1 (expressed in terms of relative proportions correct at different serial positions) is obtained. Thus the same curve is predicted for all three conditions, because the actual frequencies of the tones make no difference to performance provided that ordinal judgments can be performed accurately. Similarly, orderly transposition gradients are predicted by the guessing model (Figure 7b). The model has no free parameters. Although the absolute performance of participants in Experiment 1 is higher than the ordinal guessing model would predict, guessing may still be contributing to the serial position curves.

Figure 7: Illustration of the predictions of the "guessing model" of absolute identification. A: Serial position curves predicted by the model; B: Response uncertainty gradients predicted by the model.

This ordinal guessing model is simple and attractive, but can be contrasted with the
SIMPLE model in the predictions it makes for Experiment 2 in which stimuli were unevenly spaced. To the extent that participants can make only ordinal judgments, then performance should be the same for the “primacy” and “recency” conditions of Experiment 2, with symmetrical serial position curves resulting in both cases. However if the judgments that participants can make are relative but better than simply ordinal, as the SIMPLE model suggests, then better performance will result on tones in the relatively sparsely-populated region of the dimension.

The results of Experiment 2 are shown in the two bottom left panels of Figure 6 (details of methodology and results are in Neath & Brown, 2005). There was no difference between the two conditions in overall performance, but there was an interaction such that in each condition items in the “crowded” region of stimulus space were less accurately identified than were items in the more sparsely populated portion of the range. More specifically: In the “recency” condition, in which higher-frequency tones were more widely spaced, the lower-frequency tones were less well identified than in the even-spacing condition. The reverse pattern occurred in the “primacy” condition.

Could the SIMPLE model fit the results from Experiment 2? We attempted to fit the results using the parameter value estimated from Experiment 1. Figure 6 shows the observed response uncertainty gradients (corrected for response bias) for both of the groups in the left-hand column, and the predictions of the model in the right-hand column. Again, the model is predicting performance reasonably well ($R^2 = 0.97$) with no new parameter estimation, although we note that the model somewhat over predicts the overall skew of the serial position curve for the “recency” tones.

In summary, the experimental data confirm that the factor influencing performance is the frequencies of the tones to be identified, not their ordinal positions. The model’s account of the data again relies on the local distinctiveness principle - items that are more widely separated relative to other nearby items will be identified relatively accurately. The results of Experiment 2 also allow us to dismiss the simple ordinal guessing model. As noted, the present model represents a simplification in several respects. These simplifications, discussed later, cause no substantial changes in the qualitative behavior of the model and its account of serial position effects.

APPLICATION OF THE MODEL TO MEMORY RETRIEVAL

We now apply SIMPLE to a range of phenomena in serial and free recall from memory, with a particular focus on serial position effects. One reason for the widespread and prolonged attention enjoyed by earlier distinctiveness models is that they formalize the intuitive idea of a link between perceptual and memory processes, such that, for example, the fact that recalling memories that are distant in time is more difficult than recalling more recent items (i.e., the recency effect) is viewed as analogous to the fact that spatially distant objects are less visually discriminable than nearby objects (Baddeley, 1976; Bjork & Whitten, 1974; Crowder, 1976; Hitch, Rejman, & Turner, 1980; Tan & Ward, 2000). The key assumption here also is that similar processes are involved in memory and identification. We have already examined the characteristic serial position curves evident in absolute identification. In the next sections we examine whether the same approach can be used to account for serial position effects in serial and free recall.

How is SIMPLE applied to memory retrieval? The central idea is the same as in
absolute identification – items from crowded regions of psychological space will be harder to retrieve because they have many close (and therefore similar) neighbors from which they must be discriminated. In the case of episodic memory, we assume (following e.g. Anderson & Matessa, 1997; Brown et al., 2000; Brown & Chater, 2001; Neath, 1993a, 1993b) that a key relevant underlying dimension is temporal. Thus items are assumed to be encoded in terms of their position along a temporal-contextual dimension. At retrieval, items’ position along a temporal distance dimension, reflecting time elapsed between learning and retrieval, governs memorability. When no other cue is available, as with lists of unrelated words, the retrieval cue may be the remembered location along this temporal dimension. Even when other retrieval cues are available, items’ position along a temporal distance dimension is assumed to be important in determining retrievability.

Thus an important factor in determining the temporal distinctiveness of items will be the temporal perspective of retrieval. At the time of retrieval, temporally distant items will be more confusable than will less temporally distant (i.e., recently encountered) items. Why is this? The temporal distance of an item’s trace is the time elapsed between item presentation and item recall. In psychological space these temporal distances are assumed to be logarithmically compressed in just the same way as are frequencies, weights, line lengths etc. The temporal distance has the same effect in memory as does the representation of a test tone’s frequency or amplitude in an absolute identification experiment: Memory traces that have similar temporal distances will be confusable and difficult to discriminate from one another.

We illustrate the case of serial recall with a figure and worked example. Figure 8a shows the temporal schedule of presentation of a five item list (A B C D E). Items are presented at a rate of one item per second, and recall is assumed to be taking place 2 s after presentation of the last item. Thus the actual temporal distances of the item’s memory representations at the time of recall range from 6 s (item A) to 2 s (item E). At encoding, items are assumed to be associated with the state of an internal temporal-contextual signal at the time the item is encountered (this is just how models such as OSCAR, Brown et al., 2000, function). In the illustration we use the simpler ratio method of calculating recall probability (cf. Equation 10), and we set $c = 1$.

![Figure 8a](image-url)

**Figure 8a:** Illustration of the application of SIMPLE to memory retrieval.

Figure 8b shows how the positions of the memories will appear from the temporal perspective of retrieval, i.e., after logarithmic transformation of the temporal distances. Note that the more temporally distant items now appear closer together, and are hence less discriminable along the retrieval

---

14 Although this is a simplification, in most experiments that we consider the experimenter has controlled for other sources of systematic variation, such as word frequency, word length, etc.
dimension. Thus in psychological space the confusability of B and C is the ratio of the smaller of their temporal distances to the larger (4/5 = 0.8), whereas the confusability of D and E is less, at 0.67. B and C will therefore be harder to discriminate from one another than will D and E. This statement in terms of the confusability/similarity of items is equivalent to the statement that B and C are closer in psychological space (the distance separating them being |log(4/5)| = 0.22) than are D and E (separation distance = 0.40).

At retrieval, participants are assumed to have access to the position of the target item’s position along the temporal dimension (Brown et al., 2000). For example, suppose item C is to be retrieved on the basis of its location. This location is illustrated by the arrow in Figure 8c. In other words, the retrieval cue is assumed to be a location in psychological space, and for present purposes the location specification is assumed to be accurate. The probability of recall, given this cue, will then be determined by how discriminable the target item is from other items, i.e., by how crowded the local temporal neighborhood is (as in the person identification analogy described earlier).

We can now calculate the probability of retrieving item C in its correct serial position using Equation 10 above. The temporal distances of the items combine to determine discriminability in just the same way as did the frequencies of tones or the lengths of lines in the absolute identification case. The recall probabilities can be calculated directly in terms of ratios of the raw temporal distances (Figure 8a). Let $TD_x$ be the temporal distance of item $x$. Thus $TD_C = 4$, and so on. Then from Equation 3 the probability of (correctly) recalling item C will be:

$$P(C | R_c) = \frac{4}{\frac{4}{6} + \frac{4}{5} + \frac{4}{4} + \frac{3}{4} + \frac{2}{4}}$$

which is 0.27, and similarly the probability of (incorrectly) recalling item D will be:

$$P(C | R_c) = \frac{3}{\frac{4}{6} + \frac{4}{5} + \frac{4}{4} + \frac{3}{4} + \frac{2}{4}}$$

which is 0.20.

The full correct serial position curve is .30, .28, .27, .28, .35, illustrating small primacy and larger recency (as typically observed in probed serial recall tasks where, as in the illustrative example, retention interval is close to the end of the list for all items). Changing the value of $c$ alters the overall level of performance, but both primacy and recency remain unless $c$ becomes small enough or large enough for performance to approach chance or ceiling levels respectively. Several key properties of the model are already evident from this example and from Figure 8. Recent items will have an advantage at retrieval after a short retention interval, because they are more locally distinctive in psychological space along the temporal dimension. Items that are near to one another in the transformed space will be more confusable with one another, and serial position effects for both primacy and recency items will occur because items near the end of the series have fewer close neighbors (an edge effect). Many of the simulations that follow show that these principles themselves suffice to account for a range of short-term and long-term episodic memory data.

The model as described above can address probed serial recall and order reconstruction tasks, because in such tasks omissions cannot occur and items must be recalled in their correct serial positions. In modeling free recall, in contrast, additional mechanisms are needed to account for omission errors and the fact that items may be recalled in response to cues for other items.
We emphasize the assumption of continuity between the absolute identification and memory retrieval cases. In both cases, items are encoded in terms of their location along some continuous dimension (e.g. frequency or amplitude in absolute identification; time in memory). In both cases, serial position curves reflect and advantage for items near the ends of the continuum. In both cases, the retrieval cue for an item’s trace is a representation of that trace’s position along the relevant continuum - this cue is provided by a test item in the absolute identification case, or, in memory retrieval, by whatever retrieval cue is internally generated or provided by the nature of the experiment. In both cases, compressive transformation of the relevant stimulus values is assumed to occur, although this happens over time, not during on-line encoding, in the memory case. To be more explicit: In the memory case the performance-limiting factor will be the remembered time of occurrence of some item relative to some reference point. The reference point will normally be the time of retrieval. Thus the application of the model to memory is consistent with the suggestion that participants are making relative, not absolute, judgments of dimensional values.

**SERIES 1: SERIAL POSITION EFFECTS IN FREE RECALL**

In the first series of simulations, we fit the model to classic serial position data from a variety of paradigms. One question of primary theoretical interest is whether a unitary model can account for serial position effects of the type that have been taken as evidence for a separate short-term store. Further groups of simulations examine forgetting and proactive interference (Series 2) and serial recall (Series 3).

*Serial Position Effects in Free Recall: Murdock (1962)*

The serial position curve characteristic of immediate free recall includes large recency and smaller primacy effects at all list lengths. Representative data are reported by Murdock (1962), who presented lists of 10, 15 or 20 items at a rate of 2 s per item for free recall. These data are shown in Figure 9, along with the output of the model. In the case of free recall, we assume that a retrieval cue for each item is based on the item’s time of occurrence relative to the time of retrieval. More specifically, the effective retrieval cue is the (log transformed) value of the time that has elapsed between the learning of that item and the time of retrieval. As with most free recall tasks, however, the precise dynamics of participants’ recall are not known, and therefore we made the simplifying assumption (discussed below) that the mean time of recall of an item was 15 s after the end of the list for the 10-item lists, and 20 s and 25 s for the 15-item and 20-item lists. This figure could in principle be set independently of the observed data; here it was not possible to do so because the relevant time interval is not known.

As the task is free recall, we need to consider two new factors. First, a list item may be recalled in response to the cue for a different list item. Thus if the list to be recalled is A B C D E, a temporal-positional recall cue for the second item might retrieve B with probability 0.6, A with probability 0.1, C with probability 0.1, and so on. (This is analogous to the absolute identification case where B might be mistakenly identified as A or C on some percentage of occasions.) In free recall, an item will be scored correct

---

15 In this, and other, experiments, we focus on the conditions of interest for the present model. Thus, although Murdock also varied presentation time per item and used other list lengths, we make no mention of these.

16 Here and elsewhere we assume that the effective retention interval for a given item is measured from the offset, not the onset, of the item presentation.
whichever cue leads to its retrieval, and therefore in modeling we took the recall probability for a given item to be the sum of its recall probabilities over all retrieval cues (subject to a maximum of 1). A second, related, issue concerns omission errors. A given retrieval cue may not cue any memory item sufficiently uniquely to enable recall of anything at all, but the one-parameter model described so far does not allow omissions. It is straightforward to do so, as follows.

Intuition suggests that low retrieval probabilities (as estimated by the model) should be most likely to lead to omissions. If retrieval probabilities are thought of as reflecting relative activation strengths, then any that fall below some threshold value will lead to omissions, and any that fall above the threshold will lead to overt recalls. Assuming some noise in activations or thresholds, this will have the overall effect of increasing recall probabilities that are already high, and reducing recall probabilities for items whose recall probabilities that are already low. We implement this via a sigmoid function such that if $P$ is an estimated recall probability from the one-parameter model used hitherto, $P$ becomes

$$P = \frac{1}{1 + e^{-\frac{t}{s}}}$$

(12)

where $t$ is the threshold and $s$ (which gives the slope of the transforming function) can be interpreted as the noisiness of the threshold. For example, if $t$ was set to 0.8 and $s$ was very large, the transformation would approximate a system that recalled (with 100% probability) all items with relative strengths greater than 0.8, and omitted (with 100% probability) all items with strengths less than 0.8. As $s$ becomes smaller, the transition from low to high recall probabilities becomes more gradual.

In application of SIMPLE to free recall there are therefore three free parameters to estimate: $c$ (temporal distinctiveness of memory representations); $t$ (threshold) and $s$ (threshold noise). In the simulation of the Murdock (1962) data, these three parameters were set to 12.9, 0.56, and 8.32 respectively for the 10-item list; 10.83, 0.42, and 12.07 (15-item list); and 9.91, 0.36, and 15.43 (20-item list).

The resulting fits (Figure 9; $R^2 = 0.97, 0.98,$ and 0.95 for the 10-, 15-, and 20-item lists) capture the key qualitative effects. If all three parameters, and the estimated retention interval, are held constant for all list lengths and re-estimated the fit remains reasonable ($R^2 = 0.92$). Both primacy and recency were obtained, although primacy was smaller in the model than in the data. We return to primacy effects below. The advantage for primacy items that is observed in the model arises because end items have fewer close neighbors than do middle-list items. The same explanation was used in explaining the serial position effects in absolute identification above. Although this edge effect also contributes to superior performance on late list items, an additional factor is required to explain the asymmetry in the

![Figure 9: Serial position effects as a function of list length in single-trial free recall. Data adapted from Murdock (1962); solid lines show performance of SIMPLE.](image-url)
curve. The large and extended recency arises because the temporal codes for items near the end of the list are more spread out and less confusible than are the codes for items earlier in the list (cf. the “primacy” and “recency” conditions of Experiment 2). This asymmetry follows directly from the logarithmic transformation carried out on the raw time values, because the transformation condenses large values more than small values. We term this “Weberian compression”, and discuss it in the context of the next simulation.

Abolition of Recency Effects after a Delay: Postman and Phillips (1965)

Early research emphasized the abolition of recency effects after a filled retention interval (Glanzer & Cunitz; 1966; Postman & Phillips, 1965); such evidence was initially seen as theoretically important in providing evidence consistent with a separate short-term store (although see Petrusic & Jamieson, 1978). Postman and Phillips had participants free recall a list of 20 items, presented at a rate of 1 per second, either immediately or after 30 seconds of interpolated activity. These data, replotted in Figure 10a, illustrate the classic large recency effect present for immediate recall but absent after a filled delay.

We examined the behavior of the model with the presentation schedule and retention intervals set as in the experiment. We assumed that the mean time of recall of an item was 20 s after the end of the experimentally-imposed retention interval. There were the same three free parameters as in the previous simulation: c (temporal distinctiveness); t (threshold) and s (threshold noise). Parameter values were chosen to fit the immediate recall condition, and then the same parameter values were applied to the delayed recall condition so that resulting differences in the serial position function could be unambiguously attributed to changes in the temporal perspective of recall rather than changes in distinctiveness or threshold.

Figure 10: Immediate (RI = 0) or delayed (RI = 30) free recall. A: Data (adapted from Postman & Phillips, 1965). B: Behavior of SIMPLE. C: Observed probability of recall as a function of position of last rehearsal. Data adapted from Rundus (1971); solid line shows performance of SIMPLE.

Figure 10b shows the behavior of the model with c = 15.1; t = 0.41, and s = 14.06. $R^2$ was 0.86 for immediate recall. Predicted recall after a delay (with the same parameter values) is evidently lower than observed; a better fit can be obtained if the threshold parameters are allowed to vary with retention interval. Most importantly, however, the model successfully predicts the abolition of
recency with the filled delay although, as before, it does not fully capture the primacy evident in the data. We deal with these effects in turn.

First, regarding recency: Crowder (1976) used the analogy of a line of telegraph poles to illustrate the ratio-like mechanism proposed by Bjork and Whitten (1974): “The items in a memory list, being presented at a constant rate, pass by with the same regularity as do telephone poles when one is on a moving train. The crucial assumption is that just as each telephone pole in the receding distance becomes less and less distinctive from its neighbors, likewise each item in the memory list becomes less distinctive from the other list items as the presentation episode recedes into the past. Therefore, retrieval probability is being assumed to depend on discriminability of traces from each other.” (Crowder, 1976, p. 462; see also Baddeley, 1976; Glenberg et al., 1983; Tan & Ward, 2000). The SIMPLE model instantiates a similar notion of discriminability. Why does the abolition of recency after a delay occur? When recall is immediate, there is less Weberian compression of the scale on which the items are represented when only a short time has elapsed since the items’ presentation. Thus there is relatively little compression on that part of the scale where late list items are represented, and substantial recency results. After a filled retention interval, in contrast, sufficient time has elapsed for the whole scale to have become compressed and so end list items lose their relative advantage almost completely. The model also predicts the observed re-emergence of recency when list presentation is slow relative to the retention interval. Bjork and Whitten (1974) found that a recency effect could be observed even after a 30 second filled retention interval if sufficient time intervened between the presentation of each list item, and suggested that “...the necessary conditions for recency-sensitive retrieval from long-term memory can be specified by an empirical law of sorts based on the ratio of the temporal separation of successive to-be-remembered items (or sets of items) to the temporal delay from those items to the point of recall” (Bjork & Whitten, 1974, p. 189). This led to the formulation of a Weber-like function relating the amount of recency to the log of the ratio of the duration of the interpresentation interval (IPI) to the retention interval (RI) (see also Baddeley, 1976; Glenberg et al., 1983; Nairne et al., 1997). Thus, if the RI increases and IPI is held constant, as in the Postman and Phillips study, the ratio becomes smaller and recency decreases. Similar behavior occurs in SIMPLE: The relative advantage of recency items decreases gradually as retention interval increases, because the spacing (and hence retrievability) of recency items relative to other list items reduces through Weberian compression as retention interval increases.

SIMPLE differs from simple ratio models in the introduction of the c parameter (so that the confusability of any two items in memory is not simply proportional to the ratio of the items’ temporal distances but is instead the ratio raised to a power) and in its emphasis on the need to discriminate memory items from several near neighbors rather than just the closest. The latter property enables SIMPLE, unlike previous ratio models (Crowder, 1976) to address primacy effects.

**Primacy Effects**

In simulating the Postman and Philips (1965) data we observed less primacy in the model than in the data. Although there is some extended primacy in the model due to edge effects, there was no significant advantage for the first few items in the model. Must primacy effects then be due partly to encoding processes, especially for longer lists? Perhaps early-list items are more strongly encoded, and the SIMPLE model
produces rather too little primacy because it does not incorporate such a factor? Crowder (1976) summarizes evidence for the contribution of active encoding processes to primacy effects. However SIMPLE makes a differential prediction from the “active encoding” account of primacy effects. Active encoding accounts must predict that primacy effects will be reduced or abolished by rehearsal-preventing activity such as continuous distraction for any list length, whereas SIMPLE predicts that primacy effects will be larger for short than for long lists and will survive even when active encoding is prevented. Consistent with this prediction, constraining participants to rehearse only the current item reduces but does not abolish primacy (Fischler, Rundus, & Atkinson, 1970; Glanzer & Meinzer, 1967; Modigliani & Hedges, 1987; Tan & Ward, 2000). More generally, primacy effects still occur when rehearsal is limited by the nature of the task (Watkins, Neath, & Sechler, 1989; Wixted & McDowell, 1989) and indeed in animals without language (Harper, McLean, & Dallymple-Alford, 1993). Rundus (1971) interpreted the evidence as consistent with the idea that other factors such as the “distinctiveness of the initial study items” are relevant (Rundus, 1971, p. 65). However SIMPLE (in contrast to both active encoding and ratio rule models) predicts that one source of primacy effects is a pure edge effect. This edge effect will, according to the model, be reduced for longer list lengths, will be evident under incidental learning conditions, and will remain when active maintenance rehearsal is prevented.

The SIMPLE model must nevertheless be able to account for the full extent of primacy under all conditions. Extant explanations refer to the importance of rehearsal and of output order and interference. Can then the SIMPLE distinctiveness approach account for the full extent of primacy in terms of edge effects in combination with the changing temporal distinctiveness produced by rehearsal and output order effects (Brody & Prytulak, 1975; Brown et al., 2000; Rundus, 1971; Tan & Ward, 2000)? Even random rehearsal will tend to “telescope” the effective temporal distances of items at the point of recall at the end of the list, such that early-list items have shorter effective retention intervals than those given by the temporal distances of the items’ initial presentations; there is ample evidence that controlled rehearsal will have this effect (e.g. Rundus, 1971; Ornstein, Naus, & Stone, 1977; Tan & Ward, 2000). The relative temporal distinctiveness of primacy items in free recall may be increased due to their relatively early recall (Murdock, 1974). In addition, the multiple and temporally distributed traces laid down by rehearsal (Modigliani & Hedges, 1987) may lead to an advantage for primacy items, for which there is opportunity for greater distribution of rehearsal, and a concomitant disadvantage for late-presented items especially after a delay (“negative recency”) due to the massed nature (or different nature: Watkins & Watkins, 1974) of rehearsals of those items (Tan & Ward, 2000). These various factors, all of which may influence the temporal distinctiveness of items at the time they are retrieved, have not yet been completely disentangled in the empirical literature, but several involve rehearsal. We examined the extent to which primacy is predicted by SIMPLE when simple recency of rehearsal is taken into account.

If time since last rehearsal is the most important factor, as the most straightforward interpretation of SIMPLE suggests, the model should be able to predict the full extent of primacy observed when free rehearsal is allowed but the probability of item recall is described as a function of last re-

---

17 This is also observed experimentally: Atkinson and Shiffrin (1971); Murdock (1962); Postman and Phillips (1965).
hearsal set. Rundus (1971) presents such data, as do Tan and Ward (2000). We averaged the data from Rundus (1971; Experiment 1) and the four relevant conditions of Tan and Ward (2000; Experiments 1 and 2). The averaged data are reproduced in Figure 10c together with the behavior of SIMPLE (c = 19.97; t = .52; s = 11.60). In obtaining this fit (R2 = .99) an average retention interval of 16 s was assumed. SIMPLE now produces the correct amount of primacy, consistent with the model’s prediction that the extent of primacy will largely be determined by the temporal distance of the last rehearsal of an item rather than the temporal distance of that item’s presentation. (Separate simulations found that the data sets of Rundus and of Tan & Ward, combined above, could be captured individually.)

Although SIMPLE predicts the extent of primacy reasonably well when the temporal distance of last rehearsal is taken into account, the simulated experimental conditions still represent an approximation to reality because differential output time effects are not taken into consideration. In free recall, it is frequently observed that the most recent items are recalled first (an adaptive strategy according to SIMPLE, for this enables them to take advantage of Weberian compression), followed by primacy items particularly if recall is immediate, followed by other items. Precise protocol of output is somewhat variable, although a strong forward bias is normally evident (e.g. Laming, 1999; Murdock, 1974). Further simulations can be conducted to estimate the effective retention intervals (defined as time elapsed between last rehearsal of an item and recall of that item) for each item that would have given rise to the observed recall probabilities. Such models, not reported in detail here, ignore the possibility of output interference but nevertheless produce a good fit to the data. This result is unsurprising, as a separate parameter is available for each item, but lends strength the claim that the SIMPLE model can give a complete account of primacy effects when the true effective retention intervals for each item are taken into consideration.

In summary: SIMPLE predicts that in free recall some primacy will occur due to the increased local distinctiveness of edge items even in the absence of rehearsal. This marks a clear difference from earlier ratio models. When rehearsal is possible, primacy effects increase due to the relative reduction in effective retention intervals for primacy items. Other factors may increase the temporal distinctiveness of early compared to midlist items; such factors may include output order or the distributed rather than massed nature of the multiple traces laid down by rehearsal or reduced encoding strength or activation for late-list items. We have not incorporated such factors into the model as they do not appear mandated by the serial position data we have considered here, although we note that such complications may be necessary to account for, e.g., negative recency effects in terms of rehearsal distribution (Modigliani & Hedges, 1987).

Apparent Departures from the Ratio Rule

The success of ratio rule models has
been seen as theoretically important because, to the extent that recall level is as predicted by a ratio rule, there is no need to postulate forgetting due to trace decay. However a challenge comes from recent claims that the absolute amount of time since item learning, not just ratios of temporal intervals, influence recall probability (Cowan, Saults, & Nugent, 1997; see also Nairne et al., 1997). This challenge is therefore also of considerable significance to the SIMPLE temporal discrimination model.

We therefore used the model to explore the data of Cowan et al. (1997) who have recently suggested that the absolute amount of time that has passed before a memory test takes place can affect auditory memory performance even when the temporal schedule of item presentations and recall is held constant in ratio terms. Such an effect would violate scale invariance and suggest a role for absolute, rather than just relative, time in forgetting. To anticipate: The SIMPLE model is used to illustrate how the Cowan et al. finding can be understood on the basis of scale-independence (and no trace decay) if the possibility of proactive interference is taken into consideration. In intuitive terms this is because the relative importance of proactive interference from earlier items increases as the IPI and RI become larger, even if the IPI and RI remain in proportion to one another.

Cowan et al. (1997) presented participants with pairs of tones, and the task was to say which tone was higher in frequency. The absolute time between the tones (retention interval) was varied. Previous studies found that performance decreased as RI increased. However Cowan et al. included a novel manipulation of the time between the first tone in the pair and the last tone of the preceding pair of tones. This scheme is illustrated in Figure 11. Each solid vertical line represents the presentation of a single tone. In Case A, a four second between-trial gap separates the second tone of Trial \( n-1 \) from the first tone on Trial \( n \), and a two second retention interval separates the first tone of Trial \( n \) from the second tone of Trial \( n \). This gives a ratio of 2:1 between the between-trial gap and the retention interval for Trial \( n \). In Case B, in contrast, each duration is 50% longer (6 s for the trial gap; 3 s for the retention interval), but the ratio remains the same: 2:1. It is therefore possible to examine the effects of varying the absolute amount of time both within and between pairs while holding the intra-pair to inter-pair temporal ratio constant.

![Figure 11: Diagrammatic illustration of the experimental procedure used by Cowan, Saults, and Nugent (1997).](image)

Cowan et al. (1997) did just this, and found that performance reduces as a function of the absolute RI even when the ratio is held constant. Their results are illustrated in Figure 12, where we show just the case where the ratio between trial gap and RI was held constant at 2:1. Cowan et al. interpreted the effect of absolute time as evidence against the suggestion that the sole determinant of performance was the ratio of the between and within pair gaps, and concluded that trace decay also played a role in forgetting (although see Cowan, Saults, & Nugent, 2001, for considerations similar to those adduced here).

However, there are two ways in which a ratio-like model such as SIMPLE, in which there is no trace decay, could accommodate such a finding. First, an absolute amount of
time is involved in the time between presentation of the second tone and the making of a response. The actual memory retrieval and comparison process must necessarily take place some finite amount of time after presentation of the second tone of the pair to be compared within a trial. This is illustrated in Figure 11, where the dashed vertical line reflects the actual time of retrieval of the first tone of a pair from memory. There is a short “response time” which is unlikely to increase in direct proportion with RI. Second, Cowan et al. (1997) explicitly assumed that the influence of previous tones on previous trials would be negligible. In Figure 11, we have included the first tone of trial \( n-1 \), which will, on average, be a constant amount of time prior to the second tone of that trial. Intuitively it is evident that the relative importance of proactive interference from these earlier trials will become greater with the passage of absolute time; indeed we use this effect below to account for forgetting in Brown-Peterson tasks.

To confirm these intuitions, we used the SIMPLE model to explore the case illustrated in Figure 11, focusing on the importance of retention interval. We examined recall of the crucial first tone in the context of just one previous pair; the RIs were set to 1.5, 3, 6, and 12 s, as gave rise to the experimental data in Figure 12, and the corresponding between-trial gaps were set to 3, 6, 12, and 24 s, thus preserving a 2:1 ratio. We set the RI for the previous (non-target) pair of items at a constant 10 s for the shortest RI, and increased response time (defined here as the time between presentation of the second tone of the target pair and the time of memory retrieval and decision-making) in proportion to the RI for the target pair. This ensured that any effect of absolute time would be due solely to the RI for the pair of tones preceding the target pair. Only the \( c \) parameter was free to vary, and with this set to 1.69 the fit shown on Figure 12 as a dashed line was obtained. A clear effect of absolute RI, of similar magnitude to that seen in the data, is observed \( (R^2 = 0.65) \). Thus the apparent effect of absolute time may be due to proactive interference from earlier trials.

Figure 12: Proportion correct performance in a tone comparison task as a function of time between tones (data adapted from Cowan, Saults, & Nugent, 1997). Filled circles show the data; dashed line illustrates performance of SIMPLE when response time is assumed fixed; solid line illustrates performance of SIMPLE when response time is free.

In a further simulation, we allowed response time to be determined independently for each RI, and all other times were scaled in proportion to retention interval. In the Cowan et al. (1997) experiment, participants were given a maximum of 2 s to make a response and so estimated response times were constrained to lie between 0.1 s and 2 s. A good fit was possible \( (R^2 = 0.92; \) see Figure 12), unsurprisingly given the number of parameters (five) in relation to the number of data points (four). More importantly, the response time estimates were in a psychologically plausible order, being 0.1 s, 0.1 s, 0.74 s, and 2.0 s for RIs of 1.5 through 12 respectively. The value of \( c \) was 1.52. Despite the overparameterization, and the fact that the
observed solution may be non-unique, the good fit indicates that incorporation of response time can allow a good account of the data to be given in a model with no trace decay. It is therefore possible that no temporal trace decay or other absolute time mechanism is needed to account for the results of Cowan et al.

Discussion

The SIMPLE local distinctiveness model appears to give a reasonable account of several key findings concerning recency effects, their abolition with filled delay, and primacy effects in free recall. It makes different predictions from other ratio-like models, in that it predicts that primacy effects in free recall can arise at retrieval. In the simulations presented so far, forgetting of a single list is entirely due to intra-list interference; there is no time-based decay of traces. Some of the evidence that has been interpreted in terms of time-based forgetting does not falsify a temporal-discrimination interference-based account of the type presented here; forgetting in the cases we have examined is best characterized as intra-list interference, because forgetting of an item occurs due to difficulty in discriminating it from its list neighbors. Thus SIMPLE explains how forgetting can occur due to the passage of time alone even though no trace decay occurs (see also Nairne, 1996).

One important prediction of SIMPLE, and one that sharply distinguishes it from the positional accounts of memory for serial order including the distinctiveness accounts of Murdock (1960) and Bower (1971) and the anchoring account of Henson (1998b), is that the form of the serial position curve should be dictated by the relative position of items along a temporal dimension, as well as (or instead of) their position along an ordinal or positional dimension. We examine this in the next simulation.

Separating Ordinal and Temporal Position: Neath and Crowder (1990)

Experiment 2 of Neath and Brown (2005) was designed to exclude the possibility that tones that differed in frequency were identified in terms of their position along an ordinal dimension (e.g. as represented by the numbers 1 through 9) rather than in terms of their position along the frequency dimension. The experiment found that the closeness of neighbors in frequency, rather than the closeness of neighbors along the positional dimension, determined performance. A similar question can be asked in the memory case. Are items retrieved in terms of their position along an ordinal or positional dimension, or their position along a temporal dimension? We have assumed the latter, but in all the experiments we have described so far items’ positions on the two dimensions are perfectly correlated, as will always be the case when presentation rate is constant. Murdock’s (1960) global distinctiveness model assumed that the positional dimension was relevant in memory (see also Bower, 1971), but Neath and Crowder (1990) assumed in contrast that temporal position was the relevant dimension. Neath and Crowder conducted an experiment in which items’ locations on the positional and temporal dimensions varied, by looking at free recall for items after presentation rate was either increased or decreased throughout the course of list presentation. We illustrate with data from two presentation schedules in their Experiment 3 (visual presentation condition). In an “increasing” schedule, five word pairs were presented (pairs visible for 2.5 s). Addition problems were presented as distracter activity for 8 seconds after the last item pair in both increasing and decreasing schedules. However in the increasing presentation schedule, the gaps between the pairs of items were filled with increasing amounts of distracter activity (0 s, 8 s, 16 s, and 32 s). In a decreasing condition, the
same gaps were used but in the reverse order. Free recall after the RI of 8 s was required in both cases. The temporal presentation schedules are therefore analogous to the “primacy” and “recency” tone sets that we used in Experiment 2. The SIMPLE model should predict relatively poor performance for the early-list items in the increasing schedule, because those items are in more crowded temporal neighborhoods. In the decreasing schedule, in contrast, performance should be reduced for late-list items but improved for early list items (compared with the increasing schedule).

The results are shown in Figure 13a and it can be seen that the predicted pattern was observed, consistent with local distinctiveness on a temporal rather than a positional dimension. Neath and Crowder found the same basic pattern in several other experiments (see also Neath & Crowder, 1996).

The SIMPLE model was given exactly the same increasing and decreasing presentation schedules as described above. We did not attempt detailed quantitative fits because pairs of items rather than single items were used in the experiment. Each pair of items was treated as a single item for the purposes of modeling. The output of the model, with $c = 0.37$ (both conditions); $t = 0.33$ and $s = 22.6$ (decreasing); and $t = 0.31$ and $s = 36.4$ (increasing), is shown in Figure 13b. It is evident that the right qualitative pattern was observed.

In summary, SIMPLE provides a reasonable qualitative account of the effects of non-constant presentation schedule on free recall probability. Put another way, the experimental data are consistent with the predictions of SIMPLE that relative temporal position, rather than just relative ordinal position, will determine how free recall varies as a function of serial position. We return to the ordinal/temporal distinction below.

**SERIES 2: INTERFERENCE-BASED FORGETTING AND ITS TIME COURSE**

The previous simulations applied SIMPLE to serial position effects in free recall, but skirted the issue of forgetting as a result of intra-seril interference and previous lists. The simulations in the present section emphasize the role of proactive interference and its interaction with temporal factors. Many recent models fail to acknowledge the fact that previous trials in an experiment can greatly affect performance on a subsequent trial. Henson (1996) found that over 40% of intrusion errors in free recall came from the
list that immediately preceded the list to be recalled (see also Estes, 1991). Furthermore, many researchers have claimed that in the absence of proactive interference (PI), little or no forgetting will occur (Keppel & Underwood, 1962; Turvey, Brick, & Osborn, 1970; Underwood, 1957). The aim of this series is to use SIMPLE to provide a perspective on PI effects, and to show that some results that have been interpreted as evidence for time-based decay rather than interference-based forgetting (e.g. Baddeley & Scott, 1971) can be reinterpreted in terms of interference.

*Forgetting as a Function of Previous Lists: Underwood (1957)*

Underwood (1957) collated the results of over a dozen published experiments, all of which examined memory performance after a 24 hour retention interval. These studies varied primarily in the number of lists that had been learned prior to the target list. The results, which are shown in Figure 14, were taken as clear evidence that "the greater the number of previous lists the greater the proactive interference" (Underwood, 1957, p. 53). When there is little or no PI, little forgetting occurs, even after 24 hours. This apparent lack of forgetting when there is no PI poses a major problem for many current models of memory.

SIMPLE can be used to address the basic data. We illustrate first with a “naive model” that treats each list as a single unit in memory. We then develop a more sophisticated account in which the amount of proactive interference for an item depends upon that item’s distinctiveness in a two-dimensional space, where the two dimensions represent item-within-list position and list-within-trial position.

The naive account illustrates the basic principles governing PI in the model. SIMPLE predicts that PI will be a major determinant of memory performance because any temporal retrieval cue for a list will be less effective when the number of lists occupying similar temporal positions to the target list increases. As a result of Weberian compression, previous lists may be quite strong candidates for retrieval on the basis of a temporal cue for a current, target, list. A current list is the most temporally distinctive because it has suffered the least amount of Weberian compression. However any temporal retrieval cue intended to retrieve that list will also act as a partial cue for the previous list and (to a lesser extent) for the list before that. Thus the major factor limiting performance will be the number of previous lists and their temporal separation from the current list. The account of PI is essentially similar to the explanation of recency effects; the same basic discrimination mechanisms are invoked in both cases.

To illustrate, we attempted to capture the basic features of the Underwood (1957) study. We assumed a 24 hour retention interval, and a 60 s gap between lists. Each list was assumed to behave in the same way as single items in the studies described above (this amounts to an assumption that the main source of PI lies in discriminating lists, rather than items within a list, from each other). The results, with $c = 1.5; t = 0.28$, and $s = 6.35$, are shown in Figure 14 (solid line, labeled “Model One”).

As can be seen, performance is at almost 100% for a single list (because there is no PI). Performance then drops off sharply as the number of previous lists increases, because each list becomes less distinctive due to the increased number of competing lists. The first additional list causes most interference, because it is the closest neighbor. Addition of more and more proactively-interfering lists has less and less additional effect, and proactive interference has reached close to its maximum level after about three or four previous lists are taken into account. This is because of the sensitiv-
ity of the model to local neighborhood — neighbors have progressively less impact on discriminability as they become more temporally distant from the to-be-discriminated item, and so neighbors more than three or four items away are sufficiently temporally distant from, and hence dissimilar to, the target list that they exert negligible influence.

Figure 14: Proportion correct performance as a function of number of previous lists. Data adapted from Underwood (1957); lines show performance of two versions of SIMPLE.

Although this basic model illustrates the mechanism of proactive interference, a more sophisticated account is needed to account for proactive interference in the case where each list contains several separate items. A complete explanation must account for the separate effects of intra-serial interference (the difficulty of discriminating a target item from other list items on the basis of a temporal retrieval cue) and inter-list interference (where forgetting of a single list is faster when that list is preceded by other, proactively interfering, lists). In particular, one important finding concerns the strong interference, during retrieval of a given item from the target list, from items that were presented, rehearsed, or recalled in the same within-list position on the previous trial during serial recall (e.g. Conrad, 1960; Estes, 1991). Can SIMPLE account for these data?

Several models of memory invoke hierarchical representations, in which items’ positions along different dimensions (e.g. item-within-list; list-within-trial) are represented simultaneously and independently (e.g. Brown et al., 2000; Burgess & Hitch, 1992, 1999; Henson, 1998b; Lee & Estes, 1981; Nairne, 1991). Within the present framework, we can conceive of items as being retrieved on the basis of their position in a two-dimensional space, where one dimension represents the within-list position of that item, and the other dimension represents the temporal position of that item within a whole set of lists. This is illustrated in Figure 15, for the case of three lists of four items. In Figure 15, the vertical axis represents within-list position (1 through 4) and the horizontal axis represents the (log transformed) amount of time between item presentation and a retrieval episode taking place immediately following the third and final list (shorter distances on the right). In SIMPLE, the retrievability of a given item will depend on its distance from its near neighbors in this two-dimensional space, in the same way as local neighborhood in a one-dimensional space (position along a simple temporal dimension) has governed performance in the simulations presented above. We assume that the distance of an item from any other item is simply the sum of its distances from that item along each of the two dimensions (i.e., we used a city block rather than Euclidean metric). The probability of identifying an item on the basis of a two-dimensional temporal retrieval cue is then an exponential function of the distance between the cue and the item’s remembered location, as before. It is evident that the discriminability of an item will depend upon the closeness of both other list items and items that occupied similar positions within other lists.
The balance between these will depend upon the gap between lists relative to the inter-item spacing within lists.

![Figure 15: Illustration of the representation of three four-item lists in two-dimensional psychological space.](image)

We examined PI in this two-dimensional model, again following the basic paradigm explored by Underwood (1957), although detailed data-fitting was not possible or appropriate given the variety of studies subsumed in the Underwood analysis. To explore the model’s performance we assumed a retention interval of 24 hours, 12-item lists assumed to be separated by 60 s (both these assumptions being unchanged from the model above), and 1 s separating each within-list item.

A fourth parameter is required in this version of the model, in order to accommodate the possibility of selective attention being paid to particular dimensions of psychological space during recall. Attentional parameters are widely used in models of categorization (see e.g., Nosofsky, 1992), and in intuitive terms can be thought of as “stretching out” the relevant psychological space in one dimension while simultaneously squashing it along another. Participants’ performance in categorization tasks is well accounted for on the assumption that they can learn to pay selective attention to the stimulus dimension that is most task relevant. We introduced a new parameter, \( w_T \), which specifies the attentional weight given to the global temporal dimension. The attention paid to the second dimension, here within-list position, is denoted \( w_p \) and is set to \((1-w_T)\) to capture the notion that increasing attention to one dimension requires a corresponding reduction in attention to others. The parameter works in the same way as equivalent parameters in models of categorization: The distance between any pair of items along each psychological dimension is multiplied by the attentional weighting parameter for that dimension before the distances along the different dimensions are summed to enter into the similarity computation.

Thus in the two-dimensional version of SIMPLE there are four free parameters: \( c, t, s, \) and \( w_T \). With these parameters set to 811, 0.39, 26.5, and 0.998 respectively, and the model applied to the Underwood (1957) data, the pattern shown in Figure 14 (dashed line, labeled “Model Two”) is obtained. Average item recall probability was used as the measure. Again it can be seen that performance drops off substantially as the amount of interference from previous lists increases.

In summary, SIMPLE shows PI, which increases as a function of the number of previous lists. PI is observed whether lists are modeled as single items or as lists. The PI arises from the reduced discriminability of items when they have increased numbers of nearby temporal neighbors which reduced the effectiveness of a given retrieval cue. We have also introduced the idea of a two-dimensional space in which items may be located, to capture the idea that items’ positions are represented both on a within-list dimension and a more global temporal dimension; we return to this theme below in the context of grouping effects.
Forgetting over Time: Peterson and Peterson (1959)

The simulations above illustrate forgetting due to PI after a retention interval that is long in relation to the time assumed to intervene between interfering and target items. Do similar principles apply in the case of short-term forgetting? Such questions are central to the scale-invariant memory assumption. The rapid forgetting over time of consonant pairs (J. Brown, 1958) or trigrams (Peterson & Peterson, 1959) when rehearsal was prevented during the retention interval, was interpreted at the time as evidence for trace decay models (J. Brown, 1958), and as evidence that different principles governed short- and long-term forgetting. If correct, this would be problematic for scale-independence claims. We therefore investigated the ability of SIMPLE to account for Brown-Peterson forgetting, using the data of Peterson and Peterson (1959). (The J. Brown, 1958, data are less suitable for modeling as a smaller number of retention intervals was used.) The data (correct recalls with latencies below 2.83 s) are reproduced in Figure 16.

We reproduced the Peterson and Peterson (1959) experimental conditions as closely as possible, treating each consonant trigram as a single item. Recall of the entire trigram was assumed to be the cube of the probability of recalling a single item (i.e., independent recalls were assumed).

In modeling the Peterson and Peterson (1959) data, a gap of 15 s between trials was assumed (as in the experiment). We assumed in the simulation that 10 previous items had been presented before the critical item, to allow for PI. Retention intervals of 1 through 18 s were used, as in the experiment, and we assumed a response time (additional to the stated retention interval) of 1 s. The results, with $c = 0.45$, $t = 0.18$, and $s = 23.3$, are shown in Figure 16. A reasonable fit was obtained ($R^2 = 0.92$). The forgetting curve produced by the model is quite well approximated by an exponential; we return to this issue later. As Laming (1992) notes, the possibility of covert rehearsals may lead to shorter effective retention intervals than those measured experimentally. Laming’s model gives a better fit under this assumption, and similarly a slightly better fit may be found with the present model if we subtracted a constant from all retention intervals.

Reduced Forgetting in the Absence of PI: Keppel and Underwood (1962)

Peterson and Peterson (1959) specifically argued against interference-based explanations of their results. However the simulation above showed that Brown-Peterson forgetting can occur as a result of interference in the SIMPLE model. This is consistent with the classic data of Keppel and Underwood (1962), who found that on the first trial of a typical short-term memory experiment, there was little or no forgetting.
over time. On the second and subsequent trials, in contrast, memory performance was worse at longer intervals and the rate at which it reduces is dependent on the number of previous trials (there is faster forgetting with a larger number of previous lists). The basic pattern of results found by Keppel and Underwood is reproduced in Figure 17a. (Note however that in some of their other studies Keppel and Underwood did find rather more forgetting on the first trial.)

![Figure 17](image)

**Figure 17:** A. Proportion correctly recalled in a Brown-Peterson task as a function of trial number and retention interval (data adapted from Keppel & Underwood, 1962). B. Behavior of the one-parameter model. C. Behavior of the three-parameter model.

This study used trigrams as stimuli; these were treated as in the previous demonstration. For a one-parameter model (not incorporating a threshold) with \( c = 1.25 \) the model’s output is shown in Figure 17b. We assumed a 60 s gap between trigrams, and that “immediate” recall took place after 1 s. The model reproduces the main features of the data well. The behavior of the model can be explained in terms of the same principles as before. First, it predicts no forgetting on the first trial, regardless of the retention interval, because there is no competition from previous list items. Second, the model predicts faster forgetting when there are more previous lists. This is because after a long retention interval, the effect of Weberian compression is such that the previous item becomes closer on the scale to the current item. This means that a temporal cue is more likely to lead to retrieval of the prior item than at shorter intervals. This effect of Weberian compression will be greater when there are more previous trials, and this is the reason for the greater rate of forgetting after a larger number of previous trials in the model. Forgetting is a little faster in the model than in the data. This may reflect the fact that recall of each of the three consonants making up a trigram is not completely independent, shorter actual than intended effective retention intervals due to covert rehearsal (cf. Laming, 1992), or the absence of a threshold in the one-parameter model. Inclusion of a threshold leads to a better fit (Figure 17c, with \( c = 0.48; t = 0.28, \) and \( s = 6.03 \)). In summary, the behavior of the model is consistent with the conclusions drawn by Keppel and Underwood (1962) themselves: Interference may provide an account of short-term Brown-Peterson forgetting in terms of the same interference mechanisms that may explain forgetting over much longer timescales.

_Further Evidence Against Trace Decay:_ Turvey, Brick, and Osborn (1970)
This explanation of PI as a fundamental source of forgetting also explains striking data reported by Turvey et al. (1970). In a modified Brown-Peterson task, Turvey et al. varied, between participants, the duration of the filled retention intervals. For the first 4 trials, one group had 10 seconds, a second group had 15 seconds, and a third group had 20 seconds of distraction prior to recall. All groups had equivalent build-up of PI by Trial 4. On the fifth trial, however, all groups counted backwards for a retention interval of 15 seconds. Of most interest is the finding that performance for the 10-15 group became worse, that of the 15-15 group remained approximately constant, and performance for the 20-15 group increased. The result is shown in Figure 18. Thus performance is determined not by retention interval (which was the same for all three groups) but by the interaction between (a) retention interval and (b) the temporal distance between the item to be recalled and proactively interfering items. These counterintuitive results represent a clear challenge to simple trace decay accounts, and have been taken as evidence for temporal trace-discriminability ratio-like accounts (e.g. Baddeley, 1976; Neath, 1998). Can SIMPLE account for them? We used the same method as used to investigate Brown-Peterson forgetting in the simulations above, although without assuming previous lists. With $c$ (the only free parameter) set to 1.92, the results shown in Figure 18 were obtained.

Why does SIMPLE behave in this way? Consider first the case where retention interval remains constant within a condition, as in the first four trials of the experiment. With a short retention interval, only a small amount of time elapses between presentation and recall, and so there is relatively little Weberian compression of the temporal location code for the to-be-remembered item.

\[ \text{Proportion Correct} \]

\[ \text{Trial} \]

Figure 18: Proportion correctly recalled in a Brown-Peterson task as a function of inter-trial spacing and retention interval (data adapted from Turvey, Brick, & Osborn, 1970). Lines show the performance of SIMPLE.

This should lead to good performance. On the other hand, the immediately prior item was learned only a short amount of time previously, and so will cause interference at retrieval. When retention interval is longer, there will be a greater amount of elapsed time and more Weberian compression of the to-be-remembered item. This should lead to worse recall for the item as compared with the shorter retention interval. Set against this, the longer retention interval also applied to the previous trial, and so there is less interference (according to the local distinctiveness principle) from previous trials than in the short-retention interval comparison. These factors act to oppose each other, with the net result that performance on the...
fourth trial of the experimental series is similar for all retention intervals. (This account differs from that of other temporal-discriminability memory models in that we assume here that all previous items enter into the calculation of memorial discriminability.)

On the fifth trial, when retention interval either increases (the 10-15 condition) or decreases (20-15 condition) the conditions separate. In the 10-15 condition, the longer retention interval worsens performance relative to the previous (fourth) trial, and the relative closeness of the interfering trials (c. 30, 20, and 10 s prior to the target item) also depresses performance. In the 20-15 condition, in contrast, less Weberian compression occurs in the 15 s retention interval than occurred over the 20 s retention interval of the previous trial, but the interfering trials are still temporally distant (60, 40, and 20 s away). Performance therefore improves on the final trial.

Trace decay models of STM have continued to receive attention despite the findings of Keppel and Underwood, Turvey et al., and others. This continued attention arises partly because of the success of trace-decay/rehearsal models in accounting for item duration effects (e.g. Baddeley, Thomson, & Buchanan, 1975; but see e.g. Brown & Hulme, 1995; Neath, Bireta, & Suprenant, 2002) and partly because of the lack of an adequate interference-based account of the relevant STM phenomena (i.e., an account of the type we are attempting to develop here). However an additional reason is the findings of Baddeley and Scott (1971; see also Marcer, 1972) that forgetting can occur following the first presentation of a list, and these findings continue to be interpreted as evidence for trace decay. Could forgetting of items within a single sequence occur due to the difficulty of temporal discrimination of the separate items within the sequence? Baddeley and Scott (1971) excluded this kind of intra-sequence explanation of their observed forgetting (Melton, 1963) on the grounds that such an account would predict faster forgetting for longer sequences and that this does not occur under appropriate comparison conditions. However in a series of simulations we found that SIMPLE does produce forgetting of a single multiple-item list, due to the difficulty of intra sequence temporal discriminability, without necessarily exhibiting faster forgetting for longer lists.

SIMPLE was given lists of 6, 9, or 12 items, each item separated by 1 s, and recall was examined for retention intervals up to 20 s. No prior sequences were presented. List recall was measured in two ways. First, the probability of correct whole-sequence recall was calculated simply by multiplying the predicted probabilities of recalling each item. With parameters that give a reasonable level of overall performance forgetting was no faster for longer lists. Faster forgetting was not obtained for longer lists even when level of initial performance was equated on the 6 item and 12 item lists by increasing c for the shorter list. Similar results are obtained with an alternative scoring method where the average recall probability was calculated. Because the interference from items is local, not global, in SIMPLE the amount of interference for a given item, being determined mainly by near neighbors, is not much influenced by additional items in a longer list. Thus according to the account we have given, forgetting of a single list of items in the absence of proactively-interfering material need not be interpreted as evidence for trace decay. Intra-sequence interference can explain single-trial forget-

22 This method is oversimplified in that it incorrectly assumes independence of recalls: Schweickert, Chen, and Poirier (1999) have shown that the probability of overall list recall is underestimated by multiplying the marginal probabilities. However the oversimplification is not crucial for the point at issue.
ting, without making incorrect predictions about the rate of forgetting for longer lists.

Release from PI: Loess and Waugh (1967); Wickens, Born and Allen, (1963)

Key evidence in favor of proactive interference explanations of short-term forgetting is the ‘release from PI’ phenomenon. When several trials of a given type (e.g. letter trigrams) are used in a Brown-Peterson paradigm, PI builds up and recall reduces across trials. Performance on a subsequent trial may improve dramatically when the nature of the material is changed (e.g. from letters to digits) and this is assumed to be due to the reduction in interference from previous trials as they are no longer similar to the target item (see, e.g., Wickens, Born, & Allen, 1963). The greater the change in the nature of the material, the greater the release from PI. The effect appears to be due to retrieval rather than encoding processes (Gardiner, Craik, & Birtwistle, 1972). Release from PI also occurs when the time interval separating successive to-be-recalled items is increased sufficiently (Loess & Waugh, 1967; Peterson & Gentile, 1965), consistent with the idea that the temporal-discrimination problem becomes easier under such circumstances (Baddeley, 1976; Baddeley & Scott, 1971).

Because there is a large literature on release from PI, we focused on SIMPLE’s ability to explain just three basic phenomena. These are (a) the release from PI after a shift in the nature of the material, (b) the dependence of the size of the release on the amount of change in the to-be-remembered material, and (c) the release from PI due to the passage of time alone. In all cases it is possible to explain the phenomena using the same mechanisms of discrimination based on distinctiveness within a local psychological neighborhood. Items will be easy to discriminate just to the extent that they are isolated from their near neighbors in psychological space. When both temporal neighborhood and semantic neighborhood are relevant, as in the classic release from PI paradigm, the discrimnability of an item in memory will depend both on its near neighbors in semantic space and on its near neighbors in temporal space, and so the model must be extended to incorporate semantic as well as temporal neighborhoods.

![Figure 19: Illustration of how release from proactive interference may result from local distinctiveness principles.](image)

In some simulations above we extended the relevant psychological space from a unidimensional time line to a two-dimensional space, where the two dimensions represented list-within-trial and item-within-list. To model release from PI, we again need to assume a two-dimensional space within which local distinctiveness is calculated, but here one dimension will represent temporal distance whereas the other will represent item similarity. A simple space is illustrated in Figure 19, for the case where three consonant trigrams are presented followed by one three-digit item.

The horizontal axis dimension represents temporal position; items are assumed to be
presented at regular temporal intervals and so in psychological space they are logarithmically spaced along this dimension. The vertical axis represents item similarity. The consonant trigrams are assumed to be similar along this dimension, and are assigned arbitrary but equal values (all items in Figure 19a; first three items in Figure 19b). The group of digits are, in contrast, assumed to have a distinct value along this dimension (see fourth item in Figure 19b), and are assigned a different value along the “semantic” dimension. In the two-dimensional space, they are therefore in much less dense local neighborhoods than the third consonant trigram, and this would be expected to lead to superior memory performance. The use of a single dimension to represent item similarity is, of course, a simplification. However it suffices for illustration, for it can be thought of as representing a conceptual quality such as ‘digit-likeness’. Whereas it would in principle be possible to construct a much richer multi-dimensional semantic space, as is typically done in exemplar models of categorization, this is not necessary for present purposes. Our point is simply that a change in item-type can result in enhanced local distinctiveness, and hence memorability, for that item even when it is part of a temporally regular sequence.

We examined the behavior of SIMPLE under the conditions described. Four trigrams were presented to the model, separated by 10 s. The first three items were given values of 1 on the “conceptual” dimension, whereas the fourth item was given a value of either 1, 1.5, or 2.5 (to represent no category shift, or a small or large category shift). The parameter $c$ was set to 1.5, and $w_T$ (the attentional weight for the temporal dimension) was 0.5. An effective retention interval of 2 s was assumed, and as before recall probability was calculated as the cube of the probability of recalling a single item. For simplicity of illustration, the no-threshold version of the model was used. The results are shown in Figure 20a, where the classic data pattern is produced: Performance reduces across the first three items, due to a build-up of PI, and then increases dramatically on the fourth item illustrating release from PI. Furthermore, the extent of the release from PI is dependent on the extent of the category shift - just as in the data (Gardiner et al., 1972; Wickens et al., 1963).

We also examined release from PI as a function of the passage of time alone (Loess & Waugh, 1967; Peterson & Gentile, 1965). As a number of authors have pointed out, release from PI after a temporal gap is difficult for many forms of classic interference theory to predict because such accounts would expect spontaneous recovery of inter-

![Figure 20: A. Release from proactive interference in SIMPLE as a function of the size of the category shift. B. Release from proactive interference in SIMPLE as a function of the passage of time.](image-url)
fering items as a function of the passage of time, and so there should be no reason to expect any time-based release from PI. We examined memory for 12 trigrams, arranged into three blocks of four items. Trials within blocks were separated by 1 s, and blocks were separated by 120 s. With $c$ set to 4, and other assumptions as in the previous simulation, the results shown in Figure 20b were obtained. Again, the right general pattern was obtained, with a gradual build up of PI when items follow in quick succession, and release from PI after a longer interval (Loess & Waugh, 1967). The explanation of the model’s behavior is similar to those that have been given before — given a temporal retrieval cue for a given item, that item will be more retrievable to the extent that it is locally distinctive in its temporal neighborhood. Items that have been closely preceded by two other items will be much less temporally distinctive than will items preceded by a large temporal gap, and so memory is better for these latter items. This general behavior of the model — time-based release from PI — is consistent with a considerable amount of evidence that PI is reduced by temporal separation both in AB-AD paradigms (Keppel, 1964; Underwood & Ekstrand, 1966; Underwood & Freund, 1968) and over shorter time periods (Alín, 1968; Kincaid & Wickens, 1970; Peterson & Gentile, 1965). Note that SIMPLE offers essentially the same explanation of time-based release from PI as was given for the ratio-rule like phenomena, and for the Turvey et al. (1970) data. The account of spontaneous recovery offered by Estes (1955) resonates well with the account here (see also Mensink & Raaijmakers, 1988).

### The Time Course of Forgetting

We are now in a position to examine the time course of forgetting in the model. This is a question of central theoretical importance; any plausible model of memory must surely have something to say on the form of, as well as the fact of, forgetting. As noted in the introduction, the possibility that the time course of forgetting may follow a power law is of particular interest in view of the scale independence of memory retrieval principles and adaptive considerations (Anderson & Milson, 1989; Anderson & Schooler, 1991). 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.

We address three related questions: (a) does the SIMPLE model necessarily exhibit precise power-law forgetting; (b) is forgetting in the model generally well characterized by a power-law or some other function, and (c) how variable is the form of the forgetting curve produced by the model? Of particular interest is the possibility that small and theoretically insignificant parametric alterations might change the form of the forgetting curve. If so, the form of the forgetting curve may not provide a naturally invariant characteristic of memory. To anticipate our conclusion: we will show that SIMPLE does not predict any simple form of forgetting curve, because the form of the forgetting curve will depend on methodological details of a given procedure and, in particular, on the extent of PI from other items. The results of exploration with SIMPLE therefore raise the possibility that the difficulty in obtaining simple universal forgetting functions, despite a century of effort (Rubin & Wenzel, 1996) may be a natural consequence of any model likely to provide an account of PI effects, and that the form of the forgetting curve does not necessarily provide a psychologically useful level of description.

First we note that despite the ratio-like properties of SIMPLE, power-law forgetting
curves need not automatically follow. Consider first a simple case where memory holds just two items. If the most recent item occurred $T$ seconds in the past, and the less recent item occurred $t$ seconds earlier, then when $c = 1$ it is straightforward to show (from equation 3) that, according to the model:

$$P = \frac{T + t}{2T + t}$$

(14)

where $P$ is the probability of recalling the more recent item (see Equation 10 above). Thus in this simple two-item case, where there is assumed to be no interference from any other items, forgetting does not follow a simple power law curve. Importantly, however, this does not exclude the possibility that when a more realistic experimental situation is modeled, a good approximation to power-law forgetting will be obtained. The form of the forgetting function seems likely to depend on the spacing of other competing items in any model that accounts for effects of the type we have modeled above.

We therefore turned to our second question, and examined the time course of forgetting in the model using the methodology described in previous demonstrations above for two-dimensional memory representations (the two dimensions being temporal distance and within-list position). First, we examined recall of a five item list, preceded by four previous lists (to allow for a realistic amount of proactive interference): 2 s separated each item within a list; 12 s separated each list; $w_T$ was 0.9. We examined forgetting for retention intervals of between 2 and 100 s (the shortest retention intervals were not used to avoid possible artifacts due to ceiling effects and the use of a retention measure with a maximum value of 1), for three values of $c$. The $t$ and $s$ parameters were held constant at 0.75 and 5 respectively. The resulting forgetting curves are shown in Figure 21, together with power-law fits. Good fits were obtained: $R^2$ values were 0.984 ($c = 2$); 0.998 ($c = 4$), and 0.992 ($c = 6$). Lower $R^2$ values were obtained for simple exponential and logarithmic functions. However we did not engage in an extensive curve-fitting exercise, as our aim was to examine whether SIMPLE would exhibit a reasonable approximation to power-law forgetting. In a series of unreported simulations we observed that a good power-law fit was also obtained for different values of $w_T$ although logarithmic curves sometimes fitted performance as well as did power-law curves.

![Figure 21: Form of the forgetting function produced by SIMPLE as a function of the $c$ parameter. Solid lines show best-fitting power law curve.](image)

The choice of performance measure is problematic in curve-fitting exercises of this type (see Rubin & Wenzel, 1996; Rubin et al., 1999; Wickens, 1999, for recent discussions). Power-law curves must predict infinite performance at zero retention intervals, but > 100% performance can not be observed on “proportion correct” performance measure. Some authors (e.g. Anderson & Schooler, 1991) therefore use “recall odds”
as the performance measure, as this is unbounded at short retention intervals. When recall odds is used as the measure of performance in SIMPLE, using the same simulated experimental conditions as described above, power-law curves fit much better than exponential or logarithmic curves even when all retention intervals from 1 s up to 250 s are used. (For example, for the data in Figure 21, the mean $R^2$ was 0.99, 0.86, and 0.73 for power, logarithmic and exponential functions respectively.) The success of power-law fits in this case is attributable partly to the power law’s particular suitability for capturing the very rapid rise in recall odds as the performance measure approaches infinity as retention interval tends toward zero. However odds measures can be unstable at low retention intervals, and reliable empirical data are difficult to obtain.

![Figure 22: Forgetting curve of SIMPLE over short time interval fit by power law (A) or exponential (B) curve.](image)

We therefore return to the use of “proportion correct” as the performance measure, as this has been most widely adopted in the empirical literature, and examine the possibility that different curves might fit SIMPLE’s forgetting performance over different timescales. Such a finding would be of some theoretical importance, for different forms of forgetting over different timescales has sometimes been taken as evidence for the operation of different memory stores. The fit of the SIMPLE model to the data obtained by Peterson and Peterson (1959; Figure 16 above), which examined forgetting over 18 s, are replotted in Figure 22. Panel A plots the output of the model on logarithmic axes; this axis transformation will lead to a power law curve appearing as a straight line. It can be seen that the behavior of the model systematically deviates from the best-fitting power-law curve. Panel B plots the same data with logarithmic transformation of the y axis only. This axis transformation will lead to an exponential curve appearing as a straight line. It is evident that an equally good fit to the model’s forgetting in Peterson and Peterson (1959) paradigm is given by an exponential curve ($R^2 = 0.99$); better fits can be obtained if the simulation is run specifically to obtain them.

Thus, taking all the results together and considering just logarithmic, exponential, and power functions, forgetting in the SIMPLE model closely follows a power law when range artifacts are avoided. However under particular circumstances forgetting in the model may be best described in terms of a power law (longer time scales; use of recall odds as performance measure); a logarithmic function (when no attention is paid to within-list distinctiveness), or an exponential function (short time scales when proportion correct is the performance measure). Further findings could be reported if we considered all the 105 functions considered by Rubin and Wenzel (1996), all of which were rejected on the basis of new data by Rubin et al. (1999). Yet exactly the same
memory model is being used in all cases. We therefore endorse the conclusion of Wickens (1999): It seems unlikely that a single form of forgetting curve will apply across different methodological circumstances. We can add a demonstration that a single relatively simple architecture can, under different circumstances, instantiate a variety of forgetting functions. An example is provided in Figure 23, which shows that forgetting (five-item list; case described above with $c = 6$) is very well characterized by an exponential curve for the first 15 s of retention ($R^2 = 0.99$) and a power law thereafter ($R^2 = 0.999$). Clearly it would be wrong to conclude from this that there are two separate stores with different characteristics that operate over different timescales.

![Figure 23: The different form of forgetting function shown by SIMPLE over short and long retention intervals.](image)

**Summary of Section Results**

What unites the simulations we have reported in this section is the emphasis on the interactive roles of proactive interference and the passage of time. One key conclusion is that the appearance of forgetting due to the passage of time alone can result from such a model as SIMPLE because the relative importance of even a single proactively-interfering item will become progressively greater as time passes due to Weberian compression of the scale on which both target and interfering items are represented. Thus the model produces some of the results that have traditionally been used to support time-based decay even though there is no decay - time based or otherwise - anywhere in the model. A second key conclusion is that many serial position effects can occur despite the absence of a separate short-term memory system.

**SERIES 3: SERIAL RECALL**

Whereas the previous simulations focus on time and proactive interference, they have had little to say about serial recall. Can the same framework be used to examine both serial and free recall? In this final set of simulations we focus on serial recall, with a specific focus on (a) serial position curves and error movement gradients in serial recall, (b) effects arising from the unfolding temporal perspective during retrieval, and (c) effects of phonemic confusability and their interaction with retention interval. The key theoretical focus remains the same: to explore the possibility that the same principles govern retrieval over many different time scales and the possibility that several short-term memory data can be explained without an assumption of trace decay.

**The Hunter-McCrary law: Bugelski (1950)**

Bugelski (1950) presented eight-item lists of nonsense syllables on a memory drum and participants were given eight trials to learn the series. Figure 24 shows the normalized proportion correct responses as a function of serial position, and the fit of the SIMPLE model. Murdock (1960) chose these data as a manifestation of data conforming to the Hunter-McCrary law (McCrary & Hunter, 1953), which we cited above as an example of scale invariance in
memory (the proportions of errors at each serial position are almost the same for many different degrees of overall learning).

![Graph showing proportion of correct responses as a function of serial position](image)

**Figure 24. Proportion of correct responses as a function of serial position from the beginning of the list (data adapted from Bugelski, 1950). Solid line shows the fit of SIMPLE.**

We assumed that participants represent items in terms of their position from the beginning of the list (because the recall task cued them to produce the first item first, the list was learned in a forward direction). Thus items are assumed to be discriminated from one another just in terms of this single dimension, and the dimensional values are assumed to be compressed logarithmically. With $c$ set to 1.34, $t = 0.25$, and $s = 22.2$, the results in Figure 24 were found in the model. Consistent with the Hunter-McCrery Law, a good fit was obtained for levels of overall performance ranging between about 20% and 60% correct. Outside this range the Hunter-McCrery Law cannot apply due to floor and ceiling effects. The simulation reflects a simplification of the experimental situation, as it ignores the fact that data points are collapsed over different degrees of learning, and it also ignores output protocol effects. Rather than increase the complexity of the current simulation, we explore output effects separately below and turn to richer and more constraining data sets.

**Error Movement Gradients I: Nairne (1992)**

A central characteristic of SIMPLE is that items near to one another in psychological space will be confusable. The tendency for systematic order errors to occur has been well documented over the past quarter of a century. In serial recall tasks and order reconstruction tasks, the same basic effect is consistently found: Items that are not recalled in their correct serial position are most likely to be recalled in a serial position adjacent to the correct one, and are progressively less likely to be recalled in a position away from the correct one as the distance between target (correct) position and recalled position increases (see, e.g., Estes, 1972; Healy, 1974, Henson, Norris, Page, & Baddeley, 1996; Nairne, 1991, 1992).

We first consider the simple case where participants are presented with a list of items in serial order, and then at recall are given the items and asked to arrange them in the order in which they were presented. In the first demonstration we ignore effects of prior lists and the time-course of recall. Figure 25 shows the results of an experiment of this type conducted by Nairne (1992). Participants were presented with five lists of five items and required to rate them for pleasantness; at test participants were provided with the items from each list, and had to place them in the order of presentation. The positional uncertainty gradients represent participants’ responses after 30 s (panel a); 4 h (panel b), and 24 h (panel c); each panel shows the different output positions into which an item was placed.

In addressing these data, we need add no new assumptions to the model. We simply calculate the probabilities of recall for every item, including the correct item. This is the
same method used in all previous simulations. The new aspect is the additional calculation of the probability of (incorrectly) recalling each item in a non-target position. Figure 25 shows the positional uncertainty gradients produced by the model for the three levels of retention interval. Following Nairne (1992), we assumed 2.5 s between items, and retention intervals of 30 s, 4 h, or 24 h. However the input to the model was simpler than that used in the experiment, for we assumed only a single five-item list (cf. five five-item lists in the experiment).

![Figure 25: Positional uncertainty gradients after retention intervals of 30 seconds, 4 hours, or 24 hours (data adapted from Nairne, 1992). Lines show the performance of SIMPLE.](image)

It can be seen that the model produces the characteristic positional uncertainty gradients as shown in Figure 25. The values of $c$ were 23.8, $3.5 \times 10^3$, and $10.4 \times 10^3$ for retention intervals of 30 s, 4 h, and 24 h respectively; there were thus just three values of one parameter for the 75 data points. The parameter values illustrate the strong tendency for longer retention intervals to be associated with higher $c$ values; we return to the issue of cross-simulation parameter values in the general discussion. As omissions are not possible in order reconstruction tasks, no threshold mechanism is needed, but any response bias was removed from the model’s output using the same correction as used in the models of absolute identification (because the experimental procedure forces each response to be produced equally often). As the temporal protocols of output are not known, we simply assumed that participants had a bias to reproduce the positions of items in forward serial order, and that each item took 5 s to place in serial position. Changing these output protocol assumptions had little effect on the $R^2$ values, which were
0.96 (30 s RI), 0.91 (4 h RI), and 0.68 (24 h RI). Thus the model gives a fair account of error movement gradients in an order reconstruction task, and the qualitative similarity of such gradients over a wide range of retention intervals. The basic effect – that the probability of placing an item close to its correct position is greater than the probability of placing an item far from its correct position – follows straightforwardly from the architecture of the model. The similarity of the memory codes for any two temporal positions falls off as a negative exponential function of the temporal distance between them, and this similarity is reflected in the movement gradients.

Figure 26. Positional uncertainty gradients for the dimensions of within-list position (A) and list- within trial (C); data calculated from Nairne (1991). B and D: Performance of SIMPLE.

One key piece of evidence for hierarchical models of memory for serial order (e.g. Brown et al., 2000; Estes, 1972; Henson 1998b) is the observation that similar error movement gradients can be seen at the level of lists within a trial as at the level of items within a list (Underwood, 1977). For example, Nairne (1991) presented five lists of five items, and asked participants to rate items for pleasantness. Two minutes later, he presented five lists of five blanks each and also the list of 25 words; participants were asked to place the words in their original list and within-list position. The characteristic uncertainty gradients were seen for both the list and the within-list dimensions. These data are shown in Figure 26a and 26c.

SIMPLE can be extended to represent the position of an item on two (or more) hierarchically arranged dimensions simultaneously. This was done for the two-dimensional model introduced in the discussion of the Underwood (1957) findings above, and the model of the release from PI. There, we assumed that items could be viewed as located in a two-dimensional space, where one dimension represented the position of an item within the trial as a whole (i.e., the set of 25 words in the Nairne, 1991, experiment) whereas the other dimension represented the position of an item within each five-item list. This scheme was illustrated in Figure 15 for shorter and fewer lists, and we applied it here to the Nairne (1991) results.

Following the Nairne (1991) methodology, in the model it was assumed that items were separated from one another by 2.5 s, and that lists were separated from one another by 5 s. A retention interval of 120 s was added in both cases. The items were represented in a two-dimensional space, with one dimension corresponding to within-list position, and one dimension corresponding to within-trial position. There were just two free parameters to predict the 625 possible data points (the probability of recall in each of the possible 25 output positions for each of the 25 items); \( c \) was set to 8.54, and \( w_T \), the attentional weight given to the within-trial dimension, was 0.91. Model response bias was removed as before. The resulting positional uncertainty gradients are shown in Figure 26b and 26d, where it is evident that the main features of the data are captured by the simple two parameter model. Nairne (1991) found a small degree of non-independence in his data (i.e., the probability that an item would be placed in the correct position on one dimension was
not independent of the probability of correct positioning on the other dimension). Additional assumptions are needed to capture this non-independence in SIMPLE. For example, if some items are not properly registered at encoding such items are likely to be placed in incorrect positions on both dimensions at recall, leading to non-independence.

This simulation has extended the scope of the model to account for the positional uncertainty gradients that are observed over two different time-scales simultaneously; this is achieved by assuming that items are located, and retrieved, in the basis of their positions in a two-dimensional neighborhood. Application of the local neighborhood rule, in exactly the same fashion as had been done throughout, results in advantages for items located in more sparsely-populated regions of psychological space. However items that are located close to one another on the within-list position dimension may have a high probability of exchange even if they are widely separated in time, because of the two-dimensional representation within which they are located.

Ordered Serial Recall: Recall Perspective Effects

Several of the simulations above have made an important simplification. The simplification concerns the fact that, in ordered serial recall tasks, the temporal perspective of recall shifts systematically during the (temporally extended) process of recall. In the SIMPLE model, shifts in recall perspective are central to the account of serial position effects in serial recall.

Serial recall of a once-presented list is a temporally extended process, in which the amount of time passing between the presentation of a given item and recall of that item will normally vary as a function of the items’ serial positions. When rate of recall is slower than rate of presentation, as is normally the case, the effective retention interval will be greater for late-list than for early-list items, at least to the extent that items are recalled in or near their correct serial positions. Indeed some recent models (e.g. Brown & Hulme, 1995; Cowan et al., 1992; Page & Norris, 1998) attribute item length effects to the additional time that passes during recall of longer items; see also Dosher (1999); Dosher and Ma (1998). Thus the possibility is raised that serial position curves in serial recall of six- or seven-item lists, which typically involve a gradual reduction in performance with increasing serial position coupled with a 5-10% increase for the very last list item, may at least in part be due to the greater effective retention interval for late-list items. This would contrast with explanations of the extended primacy effects observed in serial recall in terms of a reduced attentional encoding parameter of some kind (e.g. Brown et al., 2000; Farrell & Lewandowsky, 2002; Lewandowsky & Farrell, 2000; Lewandowsky, 1999; Lewandowsky & Murdock, 1989) or in terms of output interference (see below).

We illustrate recall perspective effects in the model by examining the effects of the time course of recall on the serial position curve produced when remembering a six-item list. Items were assumed to be presented at the rate of two per second, and to allow for the possibility of a realistic amount of PI it was assumed that four further lists preceded the target list. Between-list separation was set at 30 s. To obtain realistic levels of performance, we set \( c \) to 15 and \( w_T \) to 0.96; these parameter values were held constant in the simulations reported below and only the time course of recall was varied.

All items recalled after same RI. First we examined the serial position curve when it was assumed that every item was recalled after exactly 3 s. In this case we would expect extended recency due to the greater temporal distance of early-list items, as is typically observed in probed serial recall.
tasks. Define the effective retention interval (ERI) as the time elapsed between presentation and recall for a given item; extended recency is expected due to the greater ERI for early list items. Some primacy is predicted due to edge effects (i.e., the same explanation as the model gives for primacy effects in free recall; see earlier demonstrations). The results are shown in Figure 27 (line labeled RI = 3); it can be seen that exactly this pattern of results was obtained. The simulated conditions (constant RI) are similar to those obtaining in probed serial recall tasks, where extended recency is typically observed experimentally. Thus consideration of recall perspective effects may enable the shift towards recency to be incorporated within a single model that can also account for extended primacy in forward serial recall (see below).

The analyses illustrate a possible key role of the shifting temporal perspective of recall in determining serial position curves in ordered recall. As in the data, almost linear reductions in performance across serial positions can occur, with the exception of the last and to some extent the penultimate item. The model predicts that slower recall should lead to steeper serial position curves. We note that this version of the model does not incorporate inhibition of items after they have been recalled once (see Henson, 1998a; Houghton, 1994; Vosden & Brown 1998, for detailed discussion). Nor is output interference incorporated (Bäuml, 1998; Tulving & Arbuckle, 1963). Both of these factors may contribute to determining the form of serial position curves. Nevertheless, accounts of serial position curves in serial recall will need to refer to the shifting temporal perspective of recall during output. This is consistent with the findings of Smith (1973) who presented lists of words from 12 categories, and controlled recall order by providing category cues in different orders.

Decelerating rate of output. The simulation just described makes the assumption that recalls occur at evenly-spaced time intervals. This accords with informal observations. There are few published data on recall latency serial position effects in serial recall (but see Anderson, Bothell, Lebiere, & Matessa, 1998; Anderson & Matessa, 1997; for evidence of longer latencies for list-initial and group-initial items) but at least in free recall inter-item response times increase as recall progresses (see, e.g., Murdock & Okada, 1970; Rohrer & Wixted, 1994; Wixted & Rohrer, 1993, 1994). We therefore examined the effect of increasing the inter-item recall times non-linearly. In one simulation we assumed successive ERI values of 0.1, 1.9, 4.3, 7.1, 10.3, and 13.8 s (RI for the nth item is $n^{1.5} - 0.9$), and in another we assumed ERI values of 0.1, 3.1, 8.1, 15.1 24.1, and 35.1 s (RI for the nth item is $n^{2} - 0.9$). These increasing durations led to an even larger and more extended primacy effect, again due to greater forgetting of late list items.

The analyses illustrate a possible key role of the shifting temporal perspective of recall in determining serial position curves in ordered recall. As in the data, almost linear reductions in performance across serial positions can occur, with the exception of the last and to some extent the penultimate item. The model predicts that slower recall should lead to steeper serial position curves. We note that this version of the model does not incorporate inhibition of items after they have been recalled once (see Henson, 1998a; Houghton, 1994; Vosden & Brown 1998, for detailed discussion). Nor is output interference incorporated (Bäuml, 1998; Tulving & Arbuckle, 1963). Both of these factors may contribute to determining the form of serial position curves. Nevertheless, accounts of serial position curves in serial recall will need to refer to the shifting temporal perspective of recall during output. This is consistent with the findings of Smith (1973) who presented lists of words from 12 categories, and controlled recall order by providing category cues in different orders.

Figure 27: The effect of changing temporal recall perspective at output on the serial position curves produced by SIMPLE.
at recall and found that recall was determined much more strongly by output position than by input position (see also Beaman & Morton, 2000; Broadbent, Cooper, Frankish, & Broadbent, 1980; Cowan, Saults, Elliott, & Moreno, 2002; Tulving & Arbuckle, 1966).

It is intuitively plausible that less familiar items will be recalled relatively slowly. Thus poor memory can cause slow recall. The present demonstration of how the reverse can also occur, with slow recall causing poor memory, points to ambiguities of interpretation for many previous serial recall experiments in which speed of output was not monitored or controlled. For example, Hulme et al. (1997) found that low frequency words were recalled less well than are higher frequency words, particularly in late serial positions. Without further evidence it is unclear whether slower output, and the resulting interactions with serial position, are really a cause or a consequence of poor representations. Maylor, Vousden, and Brown (1999) found that the serial recall of elderly participants was characterized by worse performance particularly at later serial positions (see Cimbalo & Brink, 1982, for a similar result). Cowan and his colleagues (e.g. Cowan et al., 1998) have found that reduced memory span in children is correlated with inter-item pauses in recall. Hulme, Newton, Cowan, Stuart, and Brown (1999) found that gaps between items were longer during recall of nonwords than during recall of words, and attributed this to the operation of redintegrative processes during gaps. In all these cases, slower output may both be caused by, and contribute to, less temporally distinctive representations in memory. Thus the direction of causality remains unclear.

**Output delay, output interference, or reduced encoding?** The simulations above raise the possibility that the extended primacy observed in forward serial recall can be accounted for in terms of forgetting during the time-course of output. We have not incorporated output interference or reduced encoding of late-list items (e.g. via differential values of the c parameter for late-list items) because such assumptions are not necessary to account for the data we have considered. However, we believe that at present insufficient data are available to exclude an interpretation of the extended primacy partly in terms of output interference or encoding. If the recall of each item leads to some reduction in the retrievability of the items not yet recalled, this would lead to primacy extending throughout recall. Similarly, error-driven learning models might be expected to encode early-presented items more strongly than late-presented items (e.g. Farrell & Lewandowsky, 2002). Accounts of associative learning have emphasized the role of output interference rather than output delay: for example, Murdock (1963) found that slow recall of paired associates actually led to better performance than did fast recall when the number of intervening recalls was held constant. It is difficult to see how interference could provide a complete account of the primacy observed in forward serial recall, however, for Tulving and Arbuckle (1966) found greater interference from learned items than recalled items when the temporal gap between presentation and recall was held constant. If the same asymmetry held for forward serial recall over short time scales, and temporal delay during output was unimportant, extended recency would be observed in forward serial recall – the opposite to the actual pattern. Further empirical research will be needed to disentangle the effects of output interference and output delay. Such experiments are difficult methodologically, for they require that covert rehearsal and recalls during the gaps between slow recalls be prevented without any concomitant increase in interference. We note that one important source of output in-
terference in SIMPLE, unimplemented in the present account, is the reduced temporal distinctiveness that will arise from the new traces laid down by item recalls; this may exacerbate effects due to the changing temporal perspective of recall.

*Intrusions from Previous Lists*

In both serial and free recall, when different items must be recalled in each list of a series of lists, many errors are intrusions from previous lists. Murdock (1974) notes that intrusions from previous lists are more likely to come from recent rather than more distant lists — for example, an item is four times as likely to be intruded from the previous list as the one before that. (Murdock cites this as evidence for the importance of temporal factors in recall; a conclusion that is, like much of Murdock’s 1974 discussion, highly consistent with the model proposed here.) In serial recall, intrusions often occur from the recall protocol of the previous trials (see Conrad, 1960; Estes, 1991; Henson, 1996); serial position is most often preserved (see e.g. Henson, 1998b), and fewer such intrusions occur when the temporal spacing between lists is large (Henson, 1996). Estes (1991) notes that in recall of trial n there are rather few intrusions of items presented on trial n-1 but not recalled in trial n-1 (in contrast to the many intrusions on trial n that come from recall on trial n-1), consistent with the suggestion that each recall is an additional learning episode capable of producing proactive interference (Henson, 1998b).

We examined such intrusions in SIMPLE, using the account of single-list serial recall described in the previous simulation. The temporal schedule of list presentation of recall was held exactly the same; we assumed the “decelerating” output protocol according to which the first item was recalled after 0.1 s and the last item was recalled after 13.8 s. In order to obtain sufficient numbers of intrusion errors from previous lists to analyze, the parameter values were changed: $c$ was set to 2, and $w_T$ was set to 0.5. (Reducing the attentional weight on the purely temporal dimension is equivalent to increasing the attentional weight on the position-within-list dimension, as the weights must sum to 1.) With these parameter settings, just over 10% of errors were intrusions from previous lists. Of these, 55% intruded into the same position as they had occupied on the previous list in which they occurred. Most intrusions came from the immediately preceding list: of intrusion errors on trial n, 49% were items from trial n-1, 23% from trial n-2, and the remainder coming from trials n-3 or earlier.

Thus the main qualitative features observed in the data concerning intrusions from previous lists are captured, although different parameter settings give rise to different quantitative data and we did not attempt detailed quantitative fitting as suitable data sets are not available. We note that from the perspective of SIMPLE, “extra-experimental intrusions” occur via the same mechanism as intrusions from previous lists; they simply reflect proactive interference from items outwith the temporal window of the experimental environment. From this perspective, “item errors” are just higher-level order errors. At least one simplification remains: we have not distinguished previous list presentations from previous list recalls. As Estes (1991) and Henson (1996) have noted, intrusions tend to come from previous recalls rather than presentations. If it is assumed that each recall is a new learning episode, however, little understanding would be gained and much complexity would be added to the model; instead we simply interpret the “list presentations” in the model as described as being produced by accurate recalls of previous lists.

*Effects of Acoustic Confusability in Serial*
Recall: Henson et al. (1996)

The effects of phonological confusability on the short-term serial recall of verbal material have been central to arguments for a separate short-term memory system. First, several researchers have noted that the nature of memory errors changes with time, such that, for example, phonemic confusion errors may dominate at short time delays but not in the longer term (Baddeley & Ecob, 1970; Conrad, 1964, 1967; Estes, 1973; Healy, 1975). Such data have buttressed claims for the existence of a separate phonologically-based short-term store. More recently, Nairne and Kelley (1999) found that at longer time delays memory for the order of lists of confusable items may actually be better than for non-confusable items, in a reversal of the traditional detrimental effect. Can SIMPLE shed light on such data?

We focus on accounting for two basic findings related to the effects of phonemic confusability. First, and most basic, is the finding that short-term memory for the serial order of verbal items is reduced when the items are phonologically confusable, with the additional errors being predominantly movement errors. Second, we examine the interactions between retention interval and confusability effects.

In all cases, we assume that confusability effects can be accommodated in terms of an extra dimension in the psychological space within which items must be discriminated in a memory task. This is illustrated, for the case of a single list of six items, in Figure 28. The horizontal dimension represents the temporal position of items at the time list presentation is complete, as normal. The vertical dimension represents a “confusability” dimension, such that items with similar values on this dimension have similar phonological representations, whereas dissimilar items have more widely spaced values on this dimension. (It would be possible to locate items in an independently-derived multidimensional articulatory space, but both explication and implementation are easier if a single dimension represents separation in multidimensional phonological space.)

Figure 28: The representation of a six-item list along both a temporal distance and a phonological confusability dimensions. A: Dissimilar items; B: Similar items.

Figure 28a depicts the case where no items are very confusable - the items are evenly spaced during presentation, and are phonologically distant from one another. Figure 28b shows the case where the items are more phonologically similar - they are therefore distinguishable from one another primarily in terms of their position along the temporal dimension, but have similar values on the second, phonological, dimension. Simply by examining the items’ local neighborhoods in this two-dimensional space, we can see that dissimilar items will be better
remembered than similar items because the dissimilar items have fewer near neighbors.

![Figure 29: Positional recall functions for serial recall of dissimilar (A) or similar (C) items (data adapted from Henson, Norris, Page, & Baddeley, 1996). B and D: Performance of SIMPLE.](image)

To illustrate, we examined SIMPLE’s memory for a single list of six items, separated by 0.4 s, at immediate recall (implemented as a delay of 1 s before recall of the first item). To allow for the possibility of a realistic amount of proactive interference we assumed the presence of four previous lists, with each list separated by 30 s. Typical results from an experiment of this type (Henson et al., 1996) are shown in Figure 29a and 29c, which illustrate the characteristic pattern of better overall performance for non-confusable than confusable lists, with the additional movement errors for the confusable items. The simulated experimental conditions simulated correspond closely to those adopted by Henson et al.

In the model the position of items was represented in the normal way - i.e., in terms of their temporal distance from the point of recall. Because precise output timings are not known, we assumed a linear rate of output with each additional item taking 1 s to recall. These figures could in principle be set independently by measuring output recall times; in practice, good model fits were achievable under a range assumptions about the time course of output. In addition to their values on the temporal dimension, items were assigned identical or different values along a “phonological” dimension according to whether lists were composed of similar or dissimilar items as illustrated in Figure 28. Items in a “similar” list were assigned identical numbers on this dimension (one of the values 1 through 5) each item in a “dissimilar” list was randomly assigned one of the values 1 through 6 without replacement. As in previous demonstrations an attentional weight parameter, $w_T$, was used to specify the relative weight given to the temporal dimension over the phonological dimension.

The result are shown in Figure 29b and 29d. Omissions are possible and observed in serial recall and there are therefore four parameters, $c$, $s$, $t$, and $w_T$, that could not in principle be set from knowledge of the experimental conditions. A good fit to the 72 data points was obtained, with $R^2 = 0.98$ for $c = 13.4$, $s = 5.9$, $t = 0.54$, and $w_T = 0.94$. A marginally better fit can be obtained if separate parameter estimates are allowed for similar and dissimilar items.

In summary, the SIMPLE model can be extended to account for confusability effects in memory. This is done by extending the dimensionality of the space in which memory items are stored, so that one dimension represents temporal position, and the other represents degree of confusability. In principle, it would be possible to use independently-derived metrics of acoustic confusability (e.g. Miller & Nicely, 1955) to determine the positions of items in a multidimensional phonological space; according to SIMPLE
memory performance could be predicted, with only the estimation of a single attentional weight parameter, on the basis of specifications of the temporal schedules of presentation and recall and an independently-derived measure of the location of to-be-remembered stimuli in phonological (or other) psychological space.

We have focused on the simplest possible explanation of phonological confusability effects in serial recall in terms of the proximity of items’ episodic traces in a two-dimensional space. Additional parameters such as noisy output thresholds can be incorporated to account for additional factors such as the small number of observed omission errors and repetition omissions; the model behaves in plausible and predictable ways when such refinements are incorporated, exhibiting for example a tendency for more omission errors to occur towards the end of the list. However the relevant data and causal mechanisms are now quite well understood in the context of previous models (reviewed below) and so we do not repeat previous theoretical work here. It is worth noting, however, that the logarithmically-compressed temporal dimension of SIMPLE provides a way of viewing of “fill-in” errors (Page & Norris, 1998). The relevant effect is that if items are not recalled in their correct position, they tend to be preferentially recalled in the next position. For example, if target (correct) sequence recall is A B C....., and B is recalled (incorrectly) in the first output position, then A is a stronger candidate for output than C in the second output position even though both A and C are, in temporal terms, equidistant from the item that is the target for position 2 (B). In SIMPLE, this tendency emerges naturally from the logarithmic compression at the heart of the model. A and B are closer together along the compressed temporal dimension than B and C, thus A is more likely than C to be recalled in place of B, as is observed in the data. This contrasts with explanations of fill-in given by models such as OSCAR (Brown et al., 2000) or the Primacy Model (Page & Norris, 1998), in which fill-in occurs due to the greater trace strengths associated with earlier list items.

Confusability and Retention Interval: Nairne and Kelley (1999)

We noted above that the effects of confusability are greatest after short-term retention intervals, and disappear or reverse after longer retention intervals. Nairne and Kelley (1999) used an order reconstruction task to obtain a relatively pure measure of order rather than item memory, and found that (when different items were used on each trial) there was a cross-over interaction such that at short retention intervals (2 s) performance was worse for similar-sounding words, whereas at longer retention intervals (24 s) order reconstruction was actually better for similar-sounding items. The interaction is shown in Figure 30 (left column). Nairne and Kelley attribute the results to the need to locate items within a multidimensional space. At short retention intervals, the factor limiting performance is the nearness of each list items to other list items. If the items are close to one another phonologically as well as positionally, the task will be more difficult. At longer retention intervals, according to the Nairne-Kelley account, the factor limiting performance changes and becomes the problem of distinguishing one list from another. Within-list similarity may therefore help performance, because items within a list are similar to each other but all different from items in neighboring lists, and so similarity can be used as a cue to aid list discrimination.
Figure 30: The reversal of the phonological similarity effect as a function of retention interval. A and C: Data (adapted from Nairne & Kelley, 1999); B and D: Performance of SIMPLE.

Nairne and Kelley (1999) do not provide a quantitative formulation of their explanation, but it is straightforward to implement a related account in SIMPLE. The account simply requires combination of features already introduced in other simulations — what is required is a multi-dimensional space, where the dimensions represent (a) within-list position, (b) within-trial position, and (c) position along a similarity dimension, as well as (d) elapsed time. At short retention intervals, there is little proactive interference from previous lists, and the performance-limiting factor is mainly the nearness of other list items (which, unlike items from previous lists, occupy similar positions along the temporal dimension). Performance will be worse on phonologically similar items (previous demonstration). As retention interval increases, items from different lists become closer to one another due to Weberian compression. As these different-list items become closer to target-list items along the position-within-trial dimension, it becomes advantageous to use an additional cue (position along the similarity dimension) to distinguish the items. We simulated recall after the longest and shortest RIs adopted by Nairne and Kelley as follows. Traces were represented along the logarithmically-transformed time-elapsed dimension in the normal way, with a 2 s RI in the immediate recall condition and a 24 s RI in the delayed recall condition (as in the Nairne & Kelley experiment). We assumed five lists of five items, to allow for the possibility of proactive interference. As in previous simulations items in a “similar” list were assigned identical numbers on the “phonological confusability” dimension (one of the values 1 through 5 was given to every item within a list) and each item in a “dissimilar” list was assigned one of the values 1 through 5. Within-list ordinal position formed the third dimension, again as in previous simulations, and the final dimension (introduced here for the first time) was “list within trial”. For phonologically confusable lists, each item was associated with a value between 1 and 5 representing the position of that item’s list within the five-list series. Thus this list-position giving value was perfectly correlated with phonological confusability value for the confusable lists. Useful values were assumed not to be available for non-confusable lists; a value of 1 on this dimension was assigned to all items. As in previous models, an attentional weight was allocated to each dimension, and these were constrained to sum to one.23

The behavior of the resulting model is illustrated in Figure 30, with c set to 8. The mean additional effective retention interval was assumed to be 5 s. SIMPLE was able to reproduce the key qualitative pattern of be-

23 Note that the absolute magnitudes of values given to items’ positions along dimensions is not important, because of the model’s ability to allocate differential attentional weights to particular dimensions.
behavior: a detrimental influence of phonological similarity on immediate recall, accompanied by a beneficial influence on delayed recall. Intuition suggests that the model achieves this by paying attention to the dimensions that are most useful for the particular RI it is faced with. Examination of the best-fit attentional weights reveals a psychological meaningful pattern. The attentional weight for the purely temporal distance dimension, \( w_T \), reduced from 0.29 (RI = 2) to 0.24 (RI = 24). The weight assigned to the list-within-trial position dimension increased (0.57 to 0.59) as did the weights on the phonological item-within-list position dimension (0.09 to 0.13). This supports the interpretation that less attention is paid to the temporal distance dimension as RI increases and values on the temporal dimension become increasingly compressed. A greater weight is given to the list-within-trial position dimension as RI increases, and this leads to an advantage for phonologically confusable items at the longer RI only. At short RIs, confusable items suffer for the reasons discussed in previous simulations. Under normal circumstances, where phonological similarity does not provide a useful cue for recall after longer retention intervals, the tendency for phonological coding over short rather than long retention intervals may partly reflect the relatively small number of phonemes in English; the small number of phonemes would lead to much proactive interference for items that are not also temporally distinctive (i.e., recent).

In summary: The SIMPLE model offers a perspective on some key data concerning item similarity effects: such effects affect predominantly within-list movement errors, and interact with retention interval. These effects can be seen as arising from a single set of retrieval principles operating over different timescales.

![Figure 31. The isolation effect for recall of mid-list items. A: Data adapted from Lippman (1980); B: Behavior of SIMPLE.](image-url)

Isolation and von Restorff effects: Lippman (1980)

Many aspects of SIMPLE’s behavior reflect the local distinctiveness of the locations that the episodic traces of items occupy in multidimensional psychological space. This general view predicts that if a single item within a list is made particularly distinctive along any dimension, then that item should be particularly memorable. This is the well-established isolation effect, or von Restorff phenomenon (see Hunt, 1995; Wallace, 1965; for reviews). Isolation effects have received a number of different interpretations, sometimes in terms of the establishment of ‘perceptual anchors’ (e.g. Lippman, 1980) or perceptual salience and differential attention (see Hunt, 1995). SIMPLE offers a
contrasting retrieval-based account.

We illustrate with data from Lippman (1980; Experiment 1). Lippman displayed a sequence of 12 CVC trigrams at a rate of 2 s per item. In the isolation condition, the seventh item was framed by a red rectangle. At test, participants were shown the 12 trigrams in random order and were required to estimate the ordinal position of each. The results, which are shown in Figure 31, were similar whether or not the seventh-presented trigram was enclosed by a red rectangle at test; Figure 31 shows only the conditions where the item was isolated at presentation but not retrieval (see Bone & Goulet, 1968; Cimbalo, Nowak, & Soderstrom, 1981, for similar results).

We addressed the data with SIMPLE using similar assumptions to those used in modeling the Bugelski (1950) data. The temporal position of items’ traces was set to the schedule of presentation. Items were also represented along a log-transformed positional dimension, and the positional cue for the mid-series item was given an increment of 10 in the condition where that item was distinctive. The probability of item recall was calculated in the normal manner. The performance of the model is shown in Figure 31b. There were two free parameters, $c$ (set to 3.25), and $w_T$, the attentional weight for the time dimension (0.12).

The key assumption is that an isolated item is distinguished from other items in terms of its position along at least one dimension. The explanation is essentially the same as has been used to account for grouping effects, phonological similarity effects, and others. Thus there is no need to assume any additional contrast-related attention or encoding devoted to the isolated item; the only difference in encoding of the isolated item relates to its different dimensional values, not its surprisingness per se (see Riefer & LaMay, 1998). Note that in the simulation above SIMPLE typically predicts a slight advantage for the items adjacent to the isolated item. Recall of items neighboring a particularly isolated item is sometimes increased, and sometimes reduced (Wallace, 1965), with memory for subsequent items being impaired when the distinctive item is particularly attention-demanding (Ellis, Determan, Runcie, McCarver, & Craig, 1971; see Fabiani & Donchin, 1995, for discussion of the possible role of encoding in the von Restorff effect).

**Grouping Effects: Hitch et al. (1996)**

A key theoretical issue for time-based models of the type proposed here is their ability to accommodate hierarchical effects (Friedman, 2001). We have already addressed the model’s ability to represent items in memory in terms of both their within-list position and their more general temporal distance from the point of recall. Other evidence for hierarchical representation in memory comes from grouping effects in memory for serial order (Frankish, 1985, 1989; Henson, 1998b; Hitch et al., 1996; Ng, 1996; Ng & Maybery, 2002; Ryan, 1969a, 1969b; Wickelgren, 1967). In a typical task participants are presented with a series of nine items, either regularly or organized into subgroups of three items, and are then required to recall all the items in correct serial order. Key findings are: (a) performance is higher overall when grouping occurs; (b) small primacy and recency effects are evident within each group, as well as at the level of the list as a whole; (c) grouping effects are larger when the structure is imposed through the insertion of temporal gaps between groups during list presentation (Ryan, 1969a); (d) auditory presentation gives rise to larger effects of grouping, but the effects discussed here are qualitatively similar for auditory and visual presentation (Frankish, 1985, 1989; Hitch et al., 1996; Ng, 1996); (e) the optimum group size is three (Wickelgren, 1967), and (f) many or-
der errors preserve within-group position (Henson, 1998b; Ryan, 1969a, 1969b).

Many of these findings fall out reasonably naturally from the SIMPLE framework. The contrasting memory representations for the two-dimensional (grouped) case and the one-dimensional (ungrouped) case are assumed to be essentially similar to those in Figure 15, although with the vertical axis representing within-group position (grouped case only) and the horizontal axis representing temporal distance from recall. The figure illustrates the idea of a trade-off — grouping causes temporally adjacent items to become more distant from each other in the two-dimensional space, and hence more memorable. Set against this, grouping can worsen performance in so far as items’ memory representations may become closer to the representations of items that are not temporally adjacent but share the same within-group position. This reduction in performance can be evident in increased numbers of errors that preserve within-group position. Does SIMPLE capture the key effects?

We examined memory for a nine-item list in the model in both a grouped and an ungrouped condition. We followed the item presentation times adopted by Hitch et al. (1996). In the model of the grouped condition, there were three groups of three items. Onsets of items within a group were separated by 0.45 s, and each group was separated by 0.5 s. In the ungrouped condition, each item onset was separated by 0.6 s. With this schedule the total time to present the list was the same in both conditions (as in the experimental methodology of Hitch et al.). On the within-group position dimension, each item was given a value of 1, 2 or 3 corresponding to its within-group position. As omissions are possible, there were four free parameters: $c$ (10.3); the attentional weight to the grouping dimension, which was assumed to be greater (0.24) in the case where the grouping dimension was highlighted by the temporal presentation schedule than in the ungrouped condition (0.06); $t$ (0.65), and $s$ (5.4). Because the time course of output is not known, we made a similar assumption as in previous simulations, i.e., that output time increased as a power function (exponent = 1.5) of the output position.

![Figure 32: Grouping effects in immediate serial recall. A: Data adapted from Hitch, Burgess, Towe, and Culpin (1996); B: Behavior of SIMPLE.](image)

The results are shown in Figure 32 for both grouped and ungrouped cases, along with the relevant conditions from Hitch et al. (1996). The basic pattern is reasonably similar to that observed experimentally when visual presentation of grouped and ungrouped lists is employed (Hitch et al., 1996): There is improved performance in the
grouped condition overall, and there are within-group serial position effects. Performance did not fall off as fast with serial position in the model as in the data; this may reflect the operation of output interference (not incorporated in the present version of SIMPLE), progressively reducing encoding for successive items at presentation, proactive interference from previous lists, or a greater slowing of response latencies as recall progresses than we have assumed.

The model’s behavior with grouped lists can be understood in terms of principles already introduced. The overall decline in performance across serial positions, for both grouped and ungrouped lists, is due to the shifting temporal perspective of recall (more time has passed when the later items are recalled). The overall advantage for items in the grouped as opposed to the ungrouped list arises because of the extra dimension on which grouped items are represented. Thus items three and four (the last item of the first group, and the first item of the second group) are close to one another, and hence not very distinctive, on the within-list temporal dimension. In the grouped case, in contrast, items three and four are distant from one another on the dimension of within-group position, and so are become more distinctive and hence discriminable within memory. This also explains the mini-primacy and recency effects that occur within each group, for the end-group items are more distinctive than are mid-group items on the position- within-group dimension. Representing items from the grouped list on the additional dimension of within-group position helps performance for the reasons given above, but also causes items that occupy the same within-group position (e.g. items four and seven, or two and five) to become closer together in psychological space. This should lead to a high proportion of order errors (exchanges) between items from the same within-group position, as is seen in the data whether presentation is auditory (Ryan, 1969a, 1969b) or visual (Henson, 1998b; Ng, 1996). We therefore examined the proportion of order errors produced by the model over different within-list separations, using the same parameters as above, and the results were as expected: Most (37%) errors were adjacent transpositions, but there were more movement errors that preserved within-group position (34%) than errors that involved movement of only two positions (24%). The precise numbers obtained vary rather substantially with parameter values however.

The nature of within-group encoding

The account of grouping given above assumes that the within-group location of items is encoded positionally rather than temporally. However the Hitch et al. (1996) experiments, in common with most grouping studies, use a constant number of items per group and regular temporal presentation of items within groups (the SOAs of items within groups remain constant). It is therefore not possible to distinguish between temporal and positional encoding of within-group position, as these are confounded. Ng and Maybery (2002) report a series of experiments designed to distinguish between (a) position coding of within-group location, and (b) encoding of within-group location in terms of the temporal distance of items from the beginning of the group. Ng and Maybery varied within-group SOA, such that for example a list of nine items might consist of an initial group of three fast-presented items (SOA = 0.5 s), followed by a group of three items presented more slowly (SOA = 1.0 s), followed by a final group of fast-presented items. Between-group order errors occur, a purely positional model of within-group encoding (according to which the timing of items within a group is not encoded on the within-group dimension) pre-
dicts under such conditions a preponderance of errors that preserve within-group position (e.g. item 5 being recalled in positions 2 or 8). A model of within-group encoding according to which items’ location within a group is represented in terms of the absolute amount of time since group onset would, in contrast, predict movement errors that preserve time since group onset (e.g. item 5 being recalled in positions 3 or 9). In a variety of experiments and conditions, Ng and Maybery found the former pattern - movement errors reflected within-group position rather than time since group onset. We examined whether SIMPLE could reproduce this effect given its assumptions of temporal distance encoding of within-list position combined with positional encoding of within-group location.

The version of the model used to model the Hitch et al. (1996) grouping data was used to address recall of a grouped lists following the temporal schedule used by Ng and Maybery (2002; Experiment 1 and 2). Because omissions were not allowed in the Ng and Maybery procedure, the model could be simplified and the slope and threshold parameters were dropped. This left two free parameters, \( c \) (estimated as 10.0) and the attentional weight to the grouping dimension (0.11). These parameter values gave good fits to the serial position curves reported by Ng and Maybery, with a grouping effect similar in magnitude to that observed in the data (an advantage of 10% for end-group over mid-group items) and a serial position curve with extended primacy (in the model performance declined from 73% in the first group to 48% in the third group; the equivalent figures were 72% and 52% in the data). More importantly, using the same error ratio statistic adopted by Ng and Maybery, the model showed many more movement errors that preserved within-group position (error ratio = 1.78 for the model; cf. 1.74 in the Ng and Maybery Experiment 1 data) than that preserved time since group onset (error ratio = 0.64 for the model; cf. 0.71 in the data). Thus SIMPLE’s assumption that the within-group location of items is positional rather than temporal is consistent with the Ng and Maybery findings, although as we discuss in more detail below such models are difficult to distinguish from those in which relative temporal position within group encodes item position (cf. Henson & Burgess, 1998).

**Temporal vs. positional encoding of order**

Is serial order information within a list coded in terms of the temporal distances of items, as emphasized here, in terms of the relative or absolute position of items within a sequence, or both? The locations of items along a positional and a temporal continuum are confounded in most experiments (when list length is constant; when rate of presentation is constant). Because the temporal distribution of items within a list influences both serial (Neath & Crowder, 1996) and free recall (Neath & Crowder, 1990), it seems clear that the location of items along a purely temporal continuum is indeed important in determining recall (see earlier model of the Neath & Crowder, 1990, results). However above we noted evidence that purely positional information is also relevant (Henson, 1999; Ng & Maybery, 2002). Such evidence also comes from studies of serial recall in which group size or list length is varied. For example, consider serial recall of a list of seven items, presented as a group of three followed by a group of four items: 1 2 3 – 4 5 6 7 (Henson, 1999). Transposition errors often preserve within-group position when group sizes are equal (Ryan, 1969,a,b). However when group sizes are unequal, movement errors might reflect the absolute time or number of positions from the start of the groups (most confusable positions 1 and 4, 2 and 5, 3 and 6); absolute time or number of positions from the end of the groups (most confusable positions 1 and
5, 2 and 6, 3 and 7), or the relative distance along either the temporal or positional range spanned by the group (most confusable positions 1 and 4, 3 and 7). Henson (1999; Experiment 1) found that transposition errors between positions 3 and 7 outnumbered errors between positions 3 and 6, and also that group-initial positions (1 and 4) were confusable. Similar effects are evident when errors across lists of different lengths are examined - for example, when a 7-item list follows a 5-item list, the fifth item recalled from the first list is likely to intrude into the seventh rather than the fifth recall position for the second list (Henson, 1999, Experiment 2).

Overall, the observed patterns of transposition errors appear to exclude the possibility that items are encoded in terms of their absolute temporal position from the start of the list. The observed pattern across lists (from one list recall to the next) appears to fit reasonably well with SIMPLE’s general assumption that at the time list presentation is complete items’ positions are encoded in term of their log-transformed relative temporal distances from the most recent end of the list. Detailed simulations are possible but complex due to the need to incorporate the time-course of recall and the possibility of output interference. We therefore focus here on the grouping results, as these appear to provide a stronger challenge to SIMPLE. The previous simulation applied SIMPLE to the results of Ng and Maybery (2002) concerning movement errors across within-list groups that varied in rate of presentation. However that version of the model, according to which the within-group location of items is encoded purely positionally, does not distinguish between encoding of (a) positions from group start; (b) positions from group end, and (c) relative within-group position. The Henson (1999) results appear to rule out (a) and perhaps also (b) (the latter on the grounds that many repetition errors occur between group-initial positions even when the groups contain unequal numbers of items). Thus the results can be used to constrain the model further.

Figure 33: Possible psychological representations of within-group and within-list position.

Figure 33 uses the same two-dimensional representation as Figure 15 to illustrate possible representations of the Henson (1999: Experiment 1) case of a group of three followed by a group of four items (constant SOA for items within a group). In all cases, the horizontal dimension represents (log transformed) temporal distance of items from the end of the list. The vertical dimension represents the location of items if encoded in terms of their position (or time) from the start of each group (panel a); their position (or time) from the end of each group (panel b), or their relative position (or relative time) through each
group (panel c). The numerals beside each filled circle represent the positions in the presented list of each item. A three-dimensional possibility, according to which position (and/or time) from both ends of the group is relevant, is not pictured. We also ignore for now the possibility that both within-group time and within-group position are represented.

Intuition suggests that the Ng and Maybery (2002) results, which suggest movement errors preserve within-group position rather than time since group onset, are consistent with the positional interpretations of (a), (b), and (c), or with the temporal interpretation of (c). Again in intuitive terms, the Henson (1999) demonstration that exchanges preserve location relative to the end of groups appear consistent with the temporal interpretation of (b), the positional interpretation of (b), and possibly the temporal or positional interpretation of (c). If movement errors simultaneously respect position from the start of the group (an issue on which the data are not yet clear when groups contain unequal numbers of items), both interpretations of (b) would additionally be excluded and the data would support either (c) or a model with additional dimensions. We confirmed these intuitions by simulation of the Henson (1999) data.

Henson (1999; Experiment One) examined serial recall of 7-item lists presented as a group of three followed by a group of four items, as described above; as a group of four followed by a group of three, or ungrouped. A response was required in each output position. The resulting serial position curves (low-confidence responses included) are reproduced in Figure 34(a). The analyses of particular interested focused on the proportion of order errors on critical positions (group-final and group-penultimate) that preserved position relative to the group’s start (start-relative errors) in comparison to the number of order errors that preserved position relative to the group’s end (end-relative errors). The main finding was that end-relative errors predominated.

![Figure 34: A. Proportion correct recall on grouped and ungrouped lists (data adapted from Henson, 1999). B. Performance of SIMPLE.](image)

We explored these findings using the model architecture developed above to model the Hitch et al. (1996) data. Simplification was possible because omissions were not permitted in the Henson (1999) procedure; the threshold and threshold-noise parameters were therefore not necessary. Within the model we varied the method of encoding within-group location, using each of the three schemes of representation illustrated in Figure 33. The temporal schedule of item presentation for the model was identical to that used experimentally, and the same assumptions about output time were made as in the previous two simulations. For each method of encoding within-group location, we obtained best-fitting values of the three free parameters ($c$; weight on the
grouping dimension for the ungrouped list, and weight on the grouping dimension for the grouped list). With all three methods of encoding within-group location, it was possible to produce a reasonable qualitative fit to the serial position curves. For example, Figure 34(b) shows serial positions obtained when position from the end of each group was encoded (panel b of Figure 33). Parameter values were 9.3 (c); 0.04 (attentional weight on the grouping dimension for ungrouped lists) and 0.08 (attentional weight on the grouping dimension for grouped lists). Better fits can be obtained if parameters are allowed to vary between the two grouping conditions (3-4 and 4-3). However the key data concern the proportion of end-relative and start-relative errors on the critical positions (the third and fourth positions within a group). Henson (1999) found that end-relative errors predominated: The experimentally observed proportion of errors on critical positions that were start-relative (at 0.13) was lower than the proportion that were end-relative (0.17). In other words, movement errors tended to be between the ends of groups rather than between the third positions of groups. Consistent with intuition, we were unable to obtain this pattern when the model encoded within-group location in terms of position from the start of the group (panel a of Figure 33) - start-relative errors predominated. However end-relative errors (0.18) predominated over start-relative errors (0.06) when within-group location was encoded in terms of distance from the end of the group (panel b of Figure 33) or when relative position within the group was encoded (panel c; 0.18 vs. 0.08). Thus the right qualitative pattern is observed, as expected, although the proportion of start-relative errors was smaller in the model than in the data (perhaps because of simplifications such as the ignoring of protrusions from previous lists). We also note that we did not analyze repetition errors in the simulation.

In summary: the data taken as a whole appear to be consistent with the core assumption of SIMPLE that the location of items within a list as a whole is encoded in terms of temporal distance. However the data do not exclude the possibility that position-within-list is also encoded. Coding of within-group position, on the other hand, cannot be represented solely in terms of absolute temporal distance or number of positions from the start of the group, or in terms of absolute temporal distance from the end of a group. A reasonable fit to the key presently available data can be given if it is assumed that within-group position is encoded in terms of the positional distance from the end of the group, or in terms of the relative time or position within a group.

**List Length Effects: Crannell and Parrish (1957)**

The final demonstration concerns list-length effects. The probability of correctly recalling a sequence of items reduces with increasing list-length according to an S-shaped function like those shown in Figure 35 for digits, letters and words (data re-calculated from Crannell & Parrish, 1957, and described in Brown et al., 2000). The characteristic high performance at short list lengths, followed by a fairly steep decline and a flattening-out as list length increases further, is accompanied by parallel functions when different materials are used.

We examined SIMPLE’s ability to account for these data. The probability of correct list length recall was assumed to be the product of the probabilities of individual item recalls. We set \( t \) to 0.52, \( s \) to 10.28, and \( c \) to 4.2, 5.4, and 7.3 for words, letters and digits respectively, and assumed items were presented at one per s. The resulting behavior of the model is shown in Figure 35 and it can be seen that a reasonably good fit was obtained. Note that the different functions for different materials were captured by
variations solely in the temporal distinctiveness parameter, \( c \).

![Figure 35: Proportion correct performance as a function of list length and item type (data adapted from Crannell & Parrish, 1957). Lines show performance of SIMPLE.](image)

The account given here represents a simplification in several respects. First, only single lists were examined in the simulation - all forgetting was due to intra-list interference, and the different values of the items on non-temporal dimensions was not taken into account. Second, the shifting perspective of recall during the temporally extended process of retrieval were not taken into account; we addressed shifting perspective effects separately above. Also, as noted above, the assumption of independence of items recalls is probably an oversimplification (Schweickert et al., 1999), and output interference may also play a role in the rate at which recall drops off for longer lists (Brown et al., 2000). Given the additional parameters involved, it is not surprising that further, more complex, simulations, not reported here, were able to reproduce the same pattern of findings when these other factors were incorporated into the model.

**GENERAL DISCUSSION**

**Summary of Findings**

We began this paper with two main claims. The first was the *scale-independent memory claim*; we suggested that similar mechanisms might govern retrieval from memory over many different time scales. The second was the *local neighborhoods claim*; we suggested that performance and serial position effects on many different identification and memory tasks might be governed by the same locally-sensitive similarity metrics as are relevant in categorization tasks if it is additionally assumed that time is an important dimension in memory. Further claims have been made: that some data may not force the assumption of trace decay in memory; that similar principles may be relevant to serial and free recall, and that the shifting temporal perspective of recall may be the basis of important memory effects. SIMPLE was applied to four types of data: absolute identification, free recall, forgetting curves, and serial recall data.

What are the general properties of SIMPLE that give rise to its qualitative behavior? Our aim has been to explain as much as possible with as few assumptions as possible, rather than to account for every nuance of the data. As applied to absolute identification and many serial recall tasks, SIMPLE has just one free parameter: \( c \). When stimuli are assumed to be represented in two dimensional space, a second attentional weight parameter, \( w_T \), must be introduced. An additional attentional weighting parameter must be introduced for each new dimension. Two further parameters are needed when omissions are made possible by the nature of the experimental task (e.g. in free recall).

What is the explanatory value of these parameters? We are keenly aware of the danger of descending to mere curve-fitting. Such concerns can be allayed in at least three ways. First, one can look for consis-
tency in parameters across simulations. We examined the best-fit values of the $c$ parameter as a function of the time-span of recall for all the simulations of serial and free recall in the present paper that were based on empirical data. It may be recalled that in modeling the near scale-invariance in absolute identification, it was assumed that the value of the $c$ parameter would scale in proportion to the range of stimuli covered. If the same account is relevant to scale-invariant effects in memory retrieval, we would expect an association between estimated $c$ values and the time over which retrieval must take place (measured as the time from list start to commencement of retrieval).

Figure 36: Estimates of the value of the c parameter as a function of the time span covered by an experiment.

The result is shown in Figure 36 ($R^2 = 0.85$ for the best-fitting power-law relation illustrated). The effect is carried partly by the three extreme points corresponding to large $c$ values; these estimates come from two separate studies (Nairne, 1992, and Underwood, 1957) but the $R^2$ reduces to 0.30 when the points are removed. Although this approach is somewhat crude, ignoring as it does all methodological differences between experiments and the value of threshold and attentional weight parameters, it provides some reassurance that the $c$ parameter is related across different simulations to important features of the task environment. A related point is that, as we have noted above, SIMPLE predicts that individual performance on memory tasks should be predictable from independent estimates of the confusability of the items along different dimensions.

Second, most of the features underpinning the SIMPLE model can be seen as independently motivated. The emphasis on temporal recency as a dimension underpinning memory retrievability has received much discussion within a rational analysis framework (e.g. Anderson & Milson, 1989; Anderson & Schooler, 1991; Brown & Chater, 2001; Brown & Vousden, 1998). The use of exponential similarity-distance metrics is independently motivated (e.g. Shepard, 1987b), and much of the machinery assumed here to underpin serial and free recall was independently developed to account for categorization data.

A third approach involves examination of the behavior of a model under parametric variation (see Li, Lewandowsky, & De-Brunner, 1996; Myung & Pitt, 1997; Roberts & Pashler, 2000). In intuitive terms, models will be preferred to the extent that they approximate the desired behavior over a wide range of parameter settings. A detailed analysis of SIMPLE along these lines will need to be the topic of a separate paper, but it is evident that basic qualitative properties of SIMPLE, such as time-based recency and its reduction with relative retention interval, or the proclivity of temporally proximal items to be recalled in the wrong

---

25 This analysis raises the possibility that $c$ should be made proportional to retention interval when modeling data from a single experiment (e.g., a variable list-length experiment). The data are at present insufficient to decide this issue.
relative order, will hold over very wide ranges of parametric variation.

Behavior of the Model

Crucial assumptions are:

1) The dimensional-cue principle. In both memory and absolute identification tasks, the effective retrieval cue for an item is assumed to be a representation of the item’s position along some continuous psychological dimension. In absolute identification tasks (e.g., identifying the position of a presented tone on the frequency dimension) the cue is the test item (the continuous dimension is frequency, or area, or weight, or some other sensory magnitude). In the case of serial or free recall from memory, an important part of the cue is the temporal distance of an item’s trace in memory (the continuous dimension is time). In this assumption the SIMPLE model has much in common with models of memory and perception proposed by Glenberg and his colleagues (Glenberg et al., 1980, 1983; Glenberg & Swanson, 1986), Murdock (1960), and Neath (e.g., Knoedler, Hellwig, & Neath, 1999; Neath, 1993a, 1993b; Neath & Crowder, 1990, 1996). As in these earlier models, the dimensional-cue principle permits the model to offer a perspective on serial position effects in both absolute identification and memory tasks.

2) The temporal memory principle. A central assumption of the SIMPLE model is that an important organizational dimension that underpins memory retrieval is temporal distance from the present (see Anderson & Schooler, 1991; Brown & Chater, 2001; Brown & Voudsen, 1998; for extensive discussion of why time provide an efficient and actual retrieval and organizational principle for memory; see Friedman, 2001, for a contrasting view). In adaptive terms, temporally proximal events may be psychologically similar and confusable in memory due to indexing by common spatio-temporal location (Gallistel, 1990) or because of the importance of temporal correlations in learning transformation-invariant object representations (e.g., Wallis, 2000).

3) The local distinctiveness principle. This principle states that the distinctiveness of an item in a memory or absolute identification task will depend on its psychological distance primarily from its nearest neighbors, rather than on its distance from every member of the list of items to be discriminated. This contrasts with assumptions embodied in current “global distinctiveness” models of memory and absolute identification (Murdock, 1960; Neath, 1993a, 1993b) and is instead akin to assumptions made in recent models of categorization (e.g., Nosofsky, 1986; Nosofsky & Palmeri, 1997).

4) The fractal memory principle. At the most general level, we assume that cognitive and perceptual systems reflect and preserve the scale-invariance and statistical self-similarity of the many fractal structures of the natural world (Chater & Brown, 1999). In the case of memory, we claim that the same processes govern memory retrieval over many different time scales. Indeed, we suggest this as a possible general psychological law; the law of scale-invariant memory (see also Brown et al., 2001; Maylor et al., 2001). We have argued that at least some of the many empirical phenomena that have previously been taken as evidence for separate short- and long-term memory systems can be explained without recourse to separate-store assumptions.

5) The Weberian Perspective principle. In common with “temporal distinctiveness” models such as that of Neath and his colleagues (see also Baddeley, 1986; Crowder, 1976; Johnson, 1991) we assume that shifts in the temporal “point of view” of recall will lead to changing perspectives on the items to

26 We note that the property of being fractal is not necessarily the same as the property of being scale-invariant or statistically self-similar.
be remembered or identified. This perspectival principle is responsible for the appearance of time-based forgetting, for the interactions of serial position effects with recall interval in the model, and for the spontaneous recovery in the recallability of information over time.

6) The interference principle. In contrast to many recent implemented and verbally-described models of memory, we assume that there is no trace decay over time. The model can nevertheless explain the appearance of time-based forgetting, due to increasing proactive interference (the main source of retrieval failure in the model) over time.

Key Theoretical Issues

Scale invariance in memory and absolute identification. Development of SIMPLE was motivated by a general hypothesis that scale invariance in the natural world will be reflected and preserved in scale invariance in the mind (Chater & Brown, 1999). To what extent does SIMPLE support the hypothesis of scale invariant memory? Here we formulate an informal distinction between weak psychological scale invariance and strong psychological scale invariance. When strong invariance obtains, it is impossible to know what the absolute level of performance is (e.g. on a memory task) from the data. An example is provided by power-law forgetting. To the extent that power-law forgetting is observed empirically (pace Rubin et al., 1999), the information that 10 words are remembered after \( t \) seconds, and 5 words are remembered after \( 2t \) seconds, would in itself give no clue as to the absolute magnitude of \( t \). The claim of weak psychological scale invariance, in contrast, concerns qualitative rather than quantitative patterns of data. An example is provided by positional uncertainty gradients (e.g. Nairne, 1992; data modeled above). Here the data change quantitatively but not qualitatively over time. Thus positional uncertainty gradients are shallower after 24 hours than they are after a few seconds, but exhibit qualitatively the same form. In modeling terms, this means that a unified explanation of the phenomena can be given over different time scales, but with different values of a performance-level determining parameter.

By these criteria, the claim most strongly sustainable in the basis of the present model and the data we have considered is one of weak scale invariance. The suggestion is that the same (qualitative) principles govern memory retrieval over many different time scales, not that there is no forgetting over time. However, SIMPLE predicts that if no other proactively interfering memories were involved, strong scale invariance would be observed empirically. Imagine an item was learned by an organism who had experienced an otherwise totally empty and uneventful life. A second item is learned a week later, and memory is tested after two more weeks. A second organism, equally devoid of previous life experience, learns one item, then learns a second item after a minute, then is tested after two minutes. We predict that the performance of the two organisms at test would be indistinguishable. In reality, of course, such an experiment could never be run. When learning events are more widely separated in time, there will always be more opportunities for additional and potentially interfering memory traces to be laid down in the interim, and performance will suffer. Thus stronger scale invariance is more likely to be seen when the to-be-remembered events are in some way distinctive from other memories, but confusable with each other. We intend our claim of scale independence in memory to be independent of any particular mechanism-level instantiation.

The claim of scale invariant memory amounts to the suggestion that the distinction between short-term and long-term
memory is not needed for the data we have considered. A detailed description of all the findings that have been taken to support an STM/LTM distinction (for comprehensive statements, see Atkinson & Shiffrin, 1968; Glanzer, 1972; Baddeley, 1976; Izawa, 1999) would be lengthy and will be the subject of a separate paper. However we emphasize that we nowhere claim that all dimensions of psychological space are equally weighted at retrieval over all time scales. For example, due to Weberian compression the position of memory traces along a temporal distance dimension is likely to be particularly helpful in distinguishing those items from their near neighbors after relatively short effective retention intervals. Thus as retention interval changes there may be a smooth and gradual shift in the attentional weights paid to different dimensions. Selective loss of the ability to represent items along some particular dimension can therefore lead to selective memory impairments. We note that arguments for separate stores are buttressed by many other methodologies that we have not considered in the present paper (e.g. McElree, 1996; Wickelgren, Corbett, & Dosher, 1980).

Relation between serial and free recall. SIMPLE attempts a perspective on serial position effects in both serial recall and free recall. According to the model, underlying memory retrieval principles are related in both cases. In both serial and free recall, the probability of successful item retrieval is governed by the local distinctiveness of items’ traces in memory. In both serial recall and free recall, an important dimension that determines distinctiveness is temporal distance. In both serial recall and free recall, shifting recall perspective due to the temporally extended process of output is responsible for many of the characteristic serial position effects. The tendency to recall recent items first in free recall is assumed to be due to the benefit such items will receive due to the compressed temporal distance dimension if and only if they are recalled early; when late-list items must be recalled after an enforced RI (either via a filled RI in free recall, or via forced forward recall in serial recall) they lose their advantage. Thus the absence of recency (of visually presented items) has the same cause in serial recall and in free recall after a filled RI. More generally, it may be possible to capture aspects of recall order in terms of the length of the path through psychological space that must be traversed to recall all the items (Romney, Brewer, & Batchelder, 1993).

The causes of forgetting. The claim made by SIMPLE is straightforward. All forgetting, over both short and long time scales, is due to interference. There is no trace decay in the model. This absence of trace decay distinguishes SIMPLE sharply from the majority of recent implemented models of both short-term and working memory (e.g. Anderson et al., 1998; Anderson & Matessa, 1997; Burgess & Hitch, 1992, 1999; Henson, 1998b; Kieras, Meyer, Mueller, & Seymour, 1999; Lovett, Reder, & Lebiere, 1999; Page & Norris, 1998; Schneider, 1999). Despite the absence of trace decay, SIMPLE can explain how forgetting may occur with the passage of time alone. This is because of the way recall perspective changes over time, and is perhaps best explained by recourse to Crowder’s (1976) telegraph pole model. When standing at the end of a line of (say) four telegraph poles, the nearby poles will appear quite distinctive. On walking away from the end of the line of poles (analogously to the passage of time) one’s perspective on the poles changes. The position of the poles along the “distance away” dimension become less distinctive (analogous to impaired item retrieval). But the telegraph poles have not moved; all that has changed is one’s perspective on them. As with telegraph poles, so with memory: As times passes, recall per-
pective on a list of items changes and the items appear less distinctive. Nothing in the items’ representations has changed; they need not have decayed or degraded in any way for forgetting to occur.

There are, of course, many different notions of decay. Most fall into two classes. The most common intuition is that “decay” must involve some change, over time, in the stored memory representations themselves. According to this type of definition, there is no trace decay in SIMPLE (although under extreme circumstances, such as head injury, physical disturbance of memory traces must always be possible at a physical level). However a second class of definition focuses on decay in memory performance rather than decay of memory representations; according to such definitions any forgetting that occurs due to the passage of time alone qualifies as decay. For example, Peterson (1966) defines decay as “forgetting which would occur no matter how dissimilar preceding and intervening activities were to the tested material” (p. 199; see also Cowan et al., 2001).

We hope that this account may shed light on a puzzle that has caused difficulty for some time. As McGeoch (1932) made clear, it makes little sense to see time in itself as a causal factor. Rather, some process correlated with the passage of time, such as rusting or erosion by acid (Posner & Konick, 1966) must be the causal agent. But such accounts have generally run into difficulties, whereas accounts based purely on interference have had difficulty in explaining apparent effects of time-based forgetting and spontaneous recovery. The concept of trace decay has proved resilient (e.g. Baddeley et al., 1975; Hebb, 1949). We hope that by showing how time-based forgetting can occur in the absence of trace decay or degradation of any kind, and by showing how the passage of time can lead to release from proactive interference (simulations above; see also Estes, 1955; Mensink & Raaijmakers, 1988) SIMPLE illustrates how key features of the data can be explained without the assumption of trace decay, spontaneous recovery of associations, consolidation, or a “Factor X” (Melton & Irwin, 1940).

SIMPLE can help understand a further paradox. Underwood and Ekstrand (1966) noted that interference theory had difficulty in explaining why forgetting is often seen on the first trial in an experiment (Keppel & Underwood, 1962; see also Peterson, 1966). A natural avenue for interference theorists is to explain away such forgetting in terms of pre-experimental interference; to the extent that this is not possible “... the decay theorist has some grounds for glee.” (Underwood & Ekstrand, 1966, p. 548). Such explanations of first-trial forgetting seemed difficult to maintain because forgetting of a list should be faster to the extent that the new list interferes with previously learned “linguistic habits”. Although this faster forgetting is not observed, SIMPLE offers an explanation of why it may not be found. According to SIMPLE, to the extent that temporal distance is a retrieval cue interference will only occur between temporally proximal items (see also Underwood & Ekstrand, 1967). Therefore, temporally distal material may not cause proactive interference even if it is similar on other dimensions. The model we have presented here can thus be seen in part as an implementation of list- differentiation theory (see Crowder, 1976).

The nature of distinctiveness in memory. The concept of “distinctiveness” is used in many different senses in the memory literature, and stands in need of rigorous definition if circularity (“better recalled memories are more distinctive”) is to be avoided. In intuitive terms, SIMPLE states that “distinctive” memories will be those that occupy relatively isolated regions of psychological space. The notion that “crowded” materials
will be remembered less well than “isolated” materials has a long history (e.g. Buxton & Newman, 1940) and has often, but not always, been taken to support some form of intraserial interference similar to the type explored in this paper (McGeoch & Irion, 1952). SIMPLE provides a formalization of the notion of “distinctiveness” that derives from categorization theory (see Appendix A for the relation to alternative formulations of distinctiveness).

The form of the forgetting function. No simple single equation governs the form of SIMPLE’s forgetting function. Despite expenditure of a considerable amount of ingenuity and empirical effort, the data appear at least consistent with the same conclusion (see e.g. Rubin & Wenzel, 1996; Rubin et al. 1999; Wickens, 1999). The forgetting curve in SIMPLE is closely approximated by exponential forgetting in the short-term, and power-law forgetting over longer time periods, but the form of the best-fitting function was found to depend to a large (and perhaps intuitively surprising) extent on parameter values that, from a theoretical point of view, seem rather peripheral to the core assumptions of the model. We therefore suggest that the search for “the” forgetting function may be misguided.

Parametric vs. nonparametric models. One of our central claims has been that many episodic memory phenomena, over both short and long time scales, can be explained using similar principles to those used in models of categorization if it is assumed that a logarithmically-compressed temporal dimension is seen as an important dimension along which memories are organized. However categorization models are often divided into two general classes, each including a wide range of specific accounts: parametric Thurstonian decision-bound models (e.g. Ashby, 1992; Ashby & Perrin, 1988), and non-parametric exemplar models (e.g. Nosofsky, 1986). Different models, though only formally equivalent under some assumptions (e.g. when a double-exponential noise distribution is assumed in Thurstonian models: Yellott, 1977) typically all do well at accounting for categorization data. Much discussion has focused on the strengths and weaknesses of the various model classes, with recent work focusing on individual participant data (Ashby, Maddox, & Lee, 1994; Maddox, 1999). Available data concerning temporal memory is not at a level of precision that permits adjudication between different model assumptions. We intend our general theoretical claims concerning scale-invariance, the absence of trace decay, and the importance of local distinctiveness along a temporal dimension to stand or fall independently of commitment to any particular model along the parametric-to-nonparametric continuum.

At the level of psychological interpretation, however, we have preferred to adopt the terminology and background assumptions of exemplar theory. There is one exception to this however. In general and intuitive terms, imperfect memory performance could be due to retrieval cue or stimulus noise, memory noise, or use of a decision rule such as the one embodied in the Luce choice model. Accounts of categorization differ in the emphasis they place on these. Models within a Thurstonian framework, such as those of Ashby and his colleagues, have emphasized the importance of perceptual noise in determining which well-defined region of psychological space a perceived item falls into. In terms of our account of episodic memory, this would translate into the assumption that temporal retrieval cues are noisy. The exemplar-based models of Nosofsky and others emphasize the role of decision processes and memory confusability in determining level of performance, and generally do not explicitly accord a major role to perceptual noise except when stimuli are highly confusable, although the need to
accommodate such noise under some circumstances is noted (Nosofsky, 1998). As Nosofsky notes, when relatively discriminable stimuli are involved the use of a “point” representation for stimuli in multidimensional space seems like a reasonable simplification. In the case of serial and free recall, the suggestion that both memory and retrieval cues are completely noise-free seems intuitively implausible. An alternative interpretation of the good fits we have obtained is that the choice-rule response selection process, with its associated free parameter (that governs the slope of the exponential function relating similarity to distance) effectively incorporates noise in the representations of memory retrieval cues (cf. Yellott, 1977). In summary: although we assume that the effects we have modeled are located at retrieval, this does not amount to the assumption that dimensional-value retrieval cues are noise-free at least in any standard Thurstonian sense.

Determinism of responding. Recent models of categorization typically allow for the possibility that responding in a classification or absolute identification task may be more of less deterministic, where the degree of determinism is given by the value of the $\gamma$ parameter in a generalization (see Ashby & Maddox, 1993) of the similarity-choice rule. When $\gamma$ has the value 1.0, responding is completely probabilistic as in the standard application of the Luce choice model. As $\gamma$ increases beyond 1.0, responding becomes increasingly deterministic. Examination of performance on simple categorization tasks often reveals values of $\gamma > 1$ for individual participants (e.g. for 24 of 28 participants in Maddox & Ashby, 1993), whereas $\gamma = 1$ for averaged data. Estes (1997b) has suggested that responding is likely to be predominantly probabilistic except when a task is sufficiently simple that good performance can be achieved by application of a simple rule.

In all the simulations of episodic memory that we have reported, good fits have been obtained with the assumption of purely probabilistic responding. This is unlikely to be due to response determinism being captured by attentional weighting parameters (Maddox, 1999) as such weighting parameters were used in a relatively small proportion of the simulations. However the $\gamma$ and c parameters serve the same function in simple versions of the model, and so their influences are not readily distinguishable. We have in any case focused on data averaged across participants, as the quality of individual participant data available in the field is not yet sufficient for model comparison. Future work will need to focus on this issue, as exemplar models of categorization, especially when they assume probabilistic responding, may be particularly successful at accounting for averaged rather than individual participant data (Ashby, Maddox, & Lee, 1994; Maddox, 1999) and SIMPLE may turn out to inherit this property.

The relation between models of episodic memory, absolute identification, and categorization. We believe that a strength of the SIMPLE model is the possibility it offers of a rapprochement between models of categorization and models of episodic memory. We have adopted much of the modeling methodology of the GCM (Nosofsky, 1986) in SIMPLE, but at the most general level our claim is that a logarithmically-compressed temporal dimension might usefully be added into any model of categorization.

However the modeling framework we have adopted is in many respects rather simpler than is typically used in modern categorization models. In particular, we have not explored the issue of response determinism at length (see above), nor have we calculated psychological distances other than by a simple city-block metric. These simplifications have been adopted largely because we were able to account for the key qualitative phenomena quite adequately without further
parameters. Further research, particularly on individual participant data, is likely to demand consideration of these additional details. Two simplifications deserve particular mention, however. The first is our use of logarithmically transformed stimulus energy or temporal distance values directly to provide the basis for both absolute identification and absolute identification; this contrasts with the derivation of a multidimensional scaling (MDS) solution used in the GCM and related models. Again, we have found that little of explanatory value is gained, and some transparency is lost, by use of MDS in analyzing the experiments above. By failing to derive MDS solutions prior to application of the similarity-choice model, however, we run the risk of failing to accommodate inter-participant variation in internal representational space. The lack of detailed individual-participant data in episodic memory renders this less of a problem than it otherwise might be. The corresponding gain, however, is that effects such as serial position curves can be attributed straightforwardly to the use of a similarity choice rule, rather than to derived psychological spaces that are ‘stretched’ at the ends. In the case of unidimensional stimuli and absolute identification, derived solutions are in practice so similar to simple logarithmically-transformed stimulus energy values that they are indistinguishable. For example, we examined the $X_i$ values (locations in psychological space) obtained by Nosofsky (1985) from Kornbrot’s (1978) data on absolute identification of loudnesses. The $X_i$ values are very well predicted by the raw stimulus (Db) values, with $R^2$ values greater than 0.995 in three out of four cases (and 0.991 for the fourth). In future work, however, it would be useful to use MDS in combination with close experimental control of temporal factors to test model predictions.

The second major simplification concerns the absence of response bias parameters. As noted above, this simplification leads to arguably more transparent accounts of the ability of other mechanisms than response bias to give rise to serial position effects, but at the cost of ignoring response bias issues in both model and data. We have aimed for explanatory simplicity and transparency at the occasional cost of detailed quantitative fit. For data examined in the present paper empirically observed response biases were in any case generally small and unsystematic and in no case did their removal significantly change the qualitative pattern of observed behavior.

Multiple traces in memory. A key theoretical claim of SIMPLE is that a separate trace is stored in memory for each episode of item occurrence (cf. also Hintzman, 1976, 1986). This distinguishes SIMPLE from several other recent models (e.g. Farrell & Lewandowsky, 2002). Indeed, according to SIMPLE it is precisely location along the temporal dimension of memory that keeps traces of items apart. There may be good adaptive reasons for preserving distinct traces of multiple episodes, for the counting of such traces is often assumed to be important in estimation and calculation (e.g. Gallistel, 1990; Gigerenzer, 2000). Many of the difficulties of “global” models of memory (Clark & Gronlund, 1996), whether applied to recognition, serial recall, or free recall, result from the agglomeration of separate episodes into a single memory or weight matrix. Crowder (1976) reviews much relevant evidence; particularly decisive points in the present context include the fact that participants can reliably distinguish separate repetitions of an item, and locate each occurrence separately (Hintzman & Block, 1971), and the fact that recency and frequency can generally although not always be distinguished (e.g. Flexser & Bower, 1974; Morton, 1968). Thus we argue that a virtue of SIMPLE is its mechanism for maintaining distinct representations of separate episodic
occurrence of items. At the same time it overcomes the intuitive implausibility of permanent availability of and access to all exemplars ever experienced. SIMPLE overcomes this problem by virtue of the dimensional locality constraint – only items that are close will be activated in the similarity-based computation; any temporally or otherwise distant items can be ignored. Thus SIMPLE offers one hypothesis of how exemplar availability may decline over time. Note that the “separate traces” issue is separate from the question of whether distinctive temporal/contextual tagging information is stored in memory – it is possible for item-to-context associations to be stored together in a single memory matrix (e.g. Brown et al., 2000) or for the same associations to be formed but stored separately (e.g. Brown, Vousden, McCormack, & Hulme, 1999). The arguments motivating SIMPLE argue for the latter possibility.

**Physical instantiation.** How could SIMPLE be instantiated at the level of mechanism? Our intention has been to abstract away from implementation at the level of nodes, connections, and weight strengths, and thus to capture the successful common features of several recent network levels while at the same time bringing models of episodic memory into alignment with previous models of categorization and identification. We therefore intend SIMPLE to stand or fall independently of particular realizations. However key features of SIMPLE can be related to mechanism-level models. As noted above, the idea that episodes of item occurrence are represented via associations of items with some form of temporal, positional, or contextual representation is present in many memory models. SIMPLE assumes some such mechanism, and emphasizes the particular importance of a purely temporal dimension (see Brown & Chater, 2001). Of more interest is the Weberian compression assumed by SIMPLE, i.e., the principle that the episodic traces of temporally distant items are more confusable than are the episodic traces of more recent items. At a mechanism level Weberian compression is assumed to arise to due to the “contextual overlap” that is assigned a central role by some models (e.g. Mensink & Raijmakers, 1988, 1989). The OSCAR model (Brown et al., 2000), for example, assumes that an oscillator-based temporal context vector changes gradually over time (e.g. during presentation of a list of to-be- recalled items). Reinstated states of this temporal context vector are used as recall probes. The temporal context vector may only be partially reinstated at recall. This leads to an advantage for recent items if they are recalled immediately after list presentation, because the non-reinstatable portion of the temporal context vector will have changed by only a small amount from its state at the time of episodic item learning (cf. also Glenberg et al., 1980). Thus the non-re reinstatable portion of the temporal context retrieval cue will act as a relatively good retrieval cue for temporally recent items, but not for temporally distant items. In OSCAR, this contextual overlap is assumed to be responsible for the extended recency observed in probed serial recall tasks and for the tendency of recent items to be output first in free recall (Deese & Kaufman, 1957; Howard & Kahana, 1999, in press; Kahana, 1996) as well as sequential dependencies in recall (Howard & Kahana, 1999). However the contextual overlap mechanism also gives rise to Weberian compression of a sort. This compression arises because from the perspective of recall, temporally distant retrieval cues will be more confusable than temporally less distant retrieval cues. Note again that the episodic associative traces in memory are not changing with the passage of time, but yet the episodic associative traces become less distinctive and discriminable with the passage of time because of
the Weberian compression arising due to temporal-contextual overlap mechanisms (see also Estes, 1950). We note the evidence given by Howard and Kahana (1999) for the importance of retrieved context (rather than simply passive context as described above) in explaining recall order effects; such accounts sit well with the model offered here.

A danger of the exemplar model approach is, for us, that it tends to lead to a particular way of thinking about memory according to which the representations of items themselves somehow reside, in recognizable form and more or less activated, in episodic memory. We do not believe this type of item-strength model is tenable (Nairne, 2001), although versions of it are widely assumed in accounts of memory over short time scales We have written rather loosely of the storage of “items”, “traces of items”, and “episodes”. To be more explicit: according to SIMPLE there is no literal sense in which “items” are stored; rather, “episodic item traces” form the content of memory. What is stored is, we suggest, the association between an item’s features and a set of features representing the temporal context of occurrence of the item.

Capacity limitations. Why is capacity limited in SIMPLE? We have suggested that a single explanation for capacity limitations in absolute identification and memory for serial order is given by SIMPLE (in contrast to Miller’s, 1956, account). In terms of the implemented model, overall level of performance (and hence channel capacity) in a given task is determined by the value of the $c$ parameter, which captures the confusability of item representations in memory. The notion that capacity limitations may arise from noise in basic neural processes has a long history (Solomons, 1900); more recently Killeen (1998) has made clear how channel capacity limitations must result from Weberian limitations – as the number of stimuli arranged along a continuum of fixed range increases, their identifiability within that context must reduce and so there is a trade-off between increasing information transmission with increasing number of items coupled with decreasing information transmission as the separation of the items decreases with their increasing number. Thus in SIMPLE capacity limitations in absolute identification and memory for serial order are assumed to arise for the same reason in each case (pace Miller, 1956).

The effectiveness of cues. In SIMPLE the effectiveness of a retrieval cue for a given item will not be determined simply by the “strength” of the item in memory, or the level of “match” between retrieval and encoding cues (see Nairne, 2001). In SIMPLE, as in the Feature Model (Nairne, 1990; Neath, 2000) the success of a retrieval cue will be determined by how similar the retrieval cue is to the cues associated with the target episode relative to the similarity of the retrieval cue to the cues associated with other traces in episodic memory. As Nairne (2001) makes clear, this prediction is very different to that made by strength-based or encoding specificity theories. In the SIMPLE model, the claim is that the distinctiveness of an item’s episodic trace is related to the proximity of other items’ episodic traces in psychological space (see also Nairne, 1988, 1990).

Serial position effects. One of our claims is that serial position effects have the same basis in absolute identification, forward serial recall, ordered serial recall, serial reconstruction, and free recall tasks. Serial position effects are assumed to arise from two factors: edge effects (reduced response competition at the ends of a stimulus range) and variable recall perspective in combination with Weberian compression of the temporal dimension along which items’ positions are assumed to be represented in episodic memory. This contrasts with previous explanations in both the absolute identifica-
tion and memory literatures (see above).

The explanation of recency effects in free recall is different to any previous account, although it has points of similarity with both “temporal-contextual overlap” accounts (Brown et al., 2000; Glenberg & Swanson, 1986) and ratio-rule accounts (Baddeley, 1976; Baddeley & Hitch, 1993; Bjork & Whitten 1974; Crowder, 1976). Detailed comparisons are in Neath and Brown (in press); here we simply note that SIMPLE contrasts with these other accounts in its ability to account for primacy effects in both serial and free recall. Primacy effects in free recall are assumed by SIMPLE to be multiply determined through the single mechanism of temporal discriminability. The temporal perspective of recall may also be important.

Relation to Other Models

The relation of SIMPLE to other models can be considered in terms of formal equivalence and in terms of the explanations given for different psychological phenomena. We have already noted the close relation between SIMPLE and exemplar models of categorization and absolute identification; here we adopt a more psychological level of analysis. There are three clear points of contrast between SIMPLE and almost all previous models of shorter-term serial recall. First, SIMPLE makes the claim that the same retrieval principles apply over all time scales. Second, SIMPLE claims that no forgetting due to trace decay occurs. Third, SIMPLE eschews the assumption of reducing attention, activation, or encoding to explain either primacy effects in forward serial recall or primacy effects more generally. We now consider specific alternative models.

Anchoring models. Anchoring models represent the values of items along some continuum in terms of the distance of each item from one or more ‘anchoring’ stimulus values. Stimuli at either end of a continuum are often assumed to act as ‘end anchors’ whereupon serial position effects are explicable in terms of mid-list items’ greater distance from end anchors and hence less precise encoding. Anchoring models have a long history of application to serial position effects in absolute identification and serial learning. Recent models of memory for serial order (Houghton 1990; 1994; Henson, 1998b; Henson et al., 1996) can be seen as types of anchoring models in that they represent the serial positions of items in terms of their association with decaying “start nodes” and “end nodes”. End-anchoring models generally do well at accounting for serial position effects, although they typically need modification to account for isolation effects (Neath et al., 2001) and experiments where list length is not known (e.g. Crowder, 1969). However it could be argued that end-anchoring models are too powerful, in that with up to four parameters to govern the strength of the start and end anchors, and two further parameters to govern the rate with which the anchor-item associative strength declines as a function of distance, it is difficult to see what form of serial position curve could not easily be accounted for by such models. The mechanism underpinning serial position effects in SIMPLE is arguably more constrained, in that local neighborhood density (which can also be a factor in end-anchoring models) is the sole source of serial position effects once SIMPLE’s parameter-free core assumption of temporal distance representation is granted. Thus SIMPLE’s explanation of serial position effects contrasts with that given by end-anchoring models.

Time-based models. The OSCAR model (Brown & Vousden, 1998; Brown et al., 1999; Brown et al., 2000; Maylor et al., 1999; Vousden & Brown, 1998; Vousden,

---

27 The number of parameters is sometimes constrained by combination (e.g. Henson, 1998b).
Brown, & Harley, 2000) and the Burgess and Hitch (Burgess, 1995; Burgess & Hitch, 1996, 1999; see also Burgess & Hitch, 1992) models assume that time-based hierarchical signals may underpin short-term serial recall. These models differ from anchoring models in that they do not assume that items’ positional codes are given directly in terms of distance from start or start and end markers. Rather, item-to-context associations are formed. The models differ in terms of their psychological interpretation and the data they address: the Burgess and Hitch model is intended as a model of the “phonological loop” component of working memory (e.g. Baddeley, 1986), whereas OSCAR in contrast is intended to apply to serial recall tasks over all time scales but unlike the Burgess and Hitch model does not directly address the data motivating the “phonological loop” account. The Burgess and Hitch model accords an important role to trace decay whereas OSCAR does not; OSCAR accords a central role to temporal-contextual overlap benefiting late-list items, whereas the Burgess and Hitch model does not. At least its most recent incarnation (Brown et al., 1999; Vousden et al., 2000) OSCAR differs from the Burgess and Hitch (1999) model in assuming that item-to-context associations are stored separately and can not interfere with one another. Thus SIMPLE differs from both models in its level of abstraction, and also differs from the Burgess and Hitch model in its denial of trace decay and in its emphasis on a logarithmically transformed time dimension. OSCAR can in many respects be seen as a mechanism-level instantiation of SIMPLE, as noted above; indeed the development of SIMPLE was motivated by the desire to capture the key explanatory elements of OSCAR within a simpler and more tractable framework. SIMPLE differs from earlier time-tagging models (Hintzman & Block, 1971) in its emphasis on temporal distance from the point of retrieval. One other model shares SIMPLE’s emphasis on the role of time in memory: the temporal distinctiveness model of Neath and his colleagues (Neath, 1993a, 1993b). SIMPLE differs from the Neath temporal distinctiveness model in its assumption that local, rather than global, temporal distinctiveness is what governs performance.

Finally, in the SIMPLE model there is no mutual exclusion between temporal and positional representations. SIMPLE assumes that the concept of distinctiveness along a temporal dimension is essential to the explanation of many key phenomena, and differs in this from recent positional models (e.g. Henson et al, 1996; Henson, 1998b). However the multidimensional psychological space assumed by SIMPLE can easily be extended to enable representation of items along a positional as well as a temporal dimension; although temporal organization is assumed to be primary we assume that psychological space will become organized along whatever dimensions are most accessible and useful for a particular task in hand. In particular, the inclusion of a non-temporal positional dimension, while not necessary to account for the data in the present paper, may be required (Henson, 1999; Ng & Maybery, 2002).

Perturbation. The perturbation model (PM) developed by Estes (1972, 1985, 1997a) emphasizes the distortion over time in memories’ attribute values over time, or over retrieval attempts. Descriptions of the model typically focus on the perturbations of items’ codes on a positional, rather than temporal, dimension; however the PM differs from SIMPLE in its emphasis on processes (per-
turbations) applying to memory representations. PM assumes that stored memories become distorted; SIMPLE emphasizes in contrast retrieval-stage interference. A further difference concerns the relation between forgetting and the passage of time. The amount of perturbation of memories’ dimensional attributes is not assumed by the PM to be a function of the passage of time alone; rather, successive retrieval attempts will increase the probability or number of perturbations (Estes, 1997a). In SIMPLE, in contrast, forgetting may occur (all other things being equal) whether or not successive retrievals have intervened. SIMPLE does not incorporate the dual-trace assumptions that form an important part of the most recent statement of the PM, and does not at present address the data that the dual-trace assumption was intended to explain. Perhaps the most fundamental point of contrast between PM and SIMPLE, however, concerns the theoretical treatment of the relation between time and other dimensions along which items are represented. In PM, perturbations are seen as additional to and separate from the multidimensional featural representations of objects in the array model of recognition and categorization (Estes, 1994). In other words, in PM there are two assumptions: (a) items are represented as vectors of features in multidimensional space, and (b) featural values may perturb over time. In SIMPLE, in contrast, time is treated as just like any other dimension; an item’s distinctiveness along a temporal distance dimension (given a particular recall perspective) affects recall in exactly the same manner as does the item’s distinctiveness along any other dimension.

Despite these large differences in the ways PM and SIMPLE are interpreted psychologically, they often make very similar predictions in practice if certain of SIMPLE’s assumptions are incorporated into the PM. For example, if the PM assumed that perturbations occurred along a logarithmically-transformed temporal distance dimension, then SIMPLE-like behavior would result for much the same reason that Thurstonian and exemplar choice model models often behave similarly. Because the PM is essentially a random walk model (with reflecting barriers corresponding to list boundaries) the probability distribution of items’ positional codes will (ignoring edge effects) tend after a fixed amount of time to approximate a normal distribution (Cox & Miller, 1968). If the original positional codes are seen as analogous to bounded regions in psychological space, then the behavior will be like that of a Thurstonian model which, as discussed above, will perform in similar fashion to an exemplar choice model.

**Feature Model.** Central to the Feature Model (FM) is a distinction between modality-dependent and modality-independent features (Nairne, 1988, 1990; Neath, 2000; Neath & Nairne, 1995; Nairne et al., 1997). This reflects the FM’s initial focus on accounting for modality effects, which have not been examined in SIMPLE. SIMPLE and the FM share the assumption that items are located in multidimensional space, although the FM uses binary features (akin to the model of Medin & Shaffer, 1978; see also Estes, 1994) whereas SIMPLE assumes continuous-valued dimensions (akin to the model of Nosofsky, 1986). Both models assume that the effectiveness of retrieval cues will depend on the extent to which they cue a given memory relative to the extent to which they cue other, competing, memories. The models differ in the assumed source of forgetting: in the FM primary memory is conceived of as a repository for retrieval cues, and forgetting occurs due to overwriting of cues rather than, as in SIMPLE, Weberian compression. However SIMPLE and the FM share the important assumption that no trace decay need be assumed.
Limitations and extensions

A full list of limitations would be extensive; we focus here on the additional mechanisms that would be needed to account for phenomena closest to those to which SIMPLE has already been applied. One issue concerns serial position effects in retrieval from semantic memory. As Healy and Parker (2001; see also Healy, Havas, & Parker, 2000) note, familiarity-based processes more complex than those envisaged in SIMPLE are involved in tasks such as reconstructing the order of US Presidents and it seems unlikely that a distinctiveness model could give a complete account of such data. Another issue concerns the strength with which items are encoded as a function of serial position. Although for the serial position data that we have considered we have not found it necessary to assumed reduced attention or encoding for successive items in a list (in contrast to other models of the same data), both intuition and empirical considerations suggest that there may be a role for such parameters in extending the model. For example, several models of human and animal learning incorporate the intuition that greater encoding will occur for items that are somehow surprising or unexpected in a given context, and the tendency for retrievability of primacy items to increase in absolute terms after a delay (see Bjork, 2001, for a review) may point to the need for inhibitory or encoding mechanisms not yet incorporated into SIMPLE. Similar encoding-level considerations arise in the context of explanations of distributed and massed practice effects (e.g. Braun & Rubin, 1998); accounts of recency-primacy shifts or distribution effects in terms of differential forgetting rates would at least in intuitive terms appear to fit less well with SIMPLE.

A further issue concerns learning and practice. In the present paper we have applied SIMPLE almost exclusively cases of single-trial learning, where one presentation of a list of non-repeated items is followed by recall. An area for future research must involve extension of the model to cases involving multiple presentations and associated learning and transfer effects. Initial exploration through simulation suggests that the retrieval assumptions of SIMPLE may combine well with assumptions of multiple-trace models such as that of Logan (1988) or Anderson, Fincham, and Douglass (1999), and that a position-from-start dimension may be important in accounting for cross-list transfer effects (see e.g. Chen, Swartz, & Terrace, 1997). However we postpone detailed consideration for future research.

There are other ways in which we believe SIMPLE could straightforwardly be extended. In most cases, the reasons we have not done so are twofold. First, our intention has been to account for as wide a range of basic effects as possible, to illustrate the scale-independence of SIMPLE’s principles of operation, while preserving as uncomplicated a framework as possible. Accounting for additional detail is possible, but only at the cost of additional parameters and mechanism. For example, it is straightforward to add additional parameters into the choice model equations to estimate guessing, background noise, etc. Such additions improve fit statistics but at the cost of increasing the number of parameters and resulting loss of transparency. It is possible to view the single c parameter as standing in for a number of other alternative parameters. Second, we believe that in many cases (e.g. the effects of response suppression and its time course; error serial position curves; modality effects; word length effects and effects of irrelevant speech and articulatory suppression) the basic mechanisms are already quite well understood, and indeed are understood sufficiently well that (a) it is straightforward to see how SIMPLE could be extended in line with previous models, to account for the relevant effects, and (b) little
real increase in psychological understanding would be achieved by extending SIMPLE to account for the effects at issue. Here, however, we briefly discuss some of these additional detailed findings.

Response suppression mechanisms, and their role in explaining errors in recall in short-term memory paradigms, are now quite well understood (e.g., Henson, 1999; Lewandowsky, 1999; Vousden & Brown, 1998). Little real explanatory gain would be achieved by incorporating such mechanisms into SIMPLE, although there would be little difficulty in doing so. Indeed we have implemented a stochastic version of the model, with response suppression included, and the essential behavior of the model is the same. Additional mechanisms are needed to explain modality effects (see Penney, 1989, for a review) and Bayesian redintegration processes along with richer multidimensional semantic representations would need to be combined with the temporal dimension described in the present paper to provide a complete account of lexicality and frequency effects in short-term memory in terms of local distinctiveness (Hulme, Maughan, & Brown, 1991).

Finally, neuropsychological evidence has been seen as central in underpinning the distinction between short-term and long-term memory. We have not addressed such data here. However we make two observations. First, systematic distortions in the locations along the temporal distance dimension can readily lead to selective preservation of more recent episodic memories in a model such as SIMPLE. Second, the claims of SIMPLE are concerned with the retrieval principles assumed to operate over all time scales and over all dimensions of psychological space. Any selective loss of the ability to represent the positions of episodic memory traces along a particular dimension can lead to selective deficits for particular types of material or for particular timescales, as can the inability to attend to a particular psychological dimension at retrieval (cf. Wickelgren, 1973).

CONCLUSIONS

We have outlined a model of scale-invariant memory and absolute identification (SIMPLE), and argued that it provides a coherent perspective on a broad range of data. SIMPLE reflects what we have argued is the scale-invariant character of perception and memory, which may in turn reflect adaptation to the scale-invariant character of many aspects of the environment. In SIMPLE, items are identified or retrieved to the degree that they are locally distinctive on the relevant dimension (whether a perceptual dimension, or, in the context of episodic memory, temporal). Distinctiveness is local because it involves comparison with neighbors of the item. Where applied to memory, SIMPLE implies that all forgetting is due to reduced local distinctiveness in psychological space, and no forgetting is due to trace decay. The same mechanisms are used in retrieval from episodic memory as are used in absolute identification and categorization tasks. Moreover, and perhaps most importantly, the same mechanisms govern retrieval over both short and long time scales for the data we have considered here, and can capture detailed regularities in data derived across time-scales.

REFERENCES


Farrell, S., & Lewandowsky, S. (2002). An e-
Estes, W.K. (1997b). Some reflections on the
Estes, W.K. (1973). Phonemic coding and r-
estes, W.K. (1972). An associative basis for
estes, W.K. (1955). Toward a statistical theory
effective distributed model of ordering in s-
Fischler, I., Rundus, D., & Atkinson, R.C.
effects in absolute judgments.
Friedman, W.J. (2001). Memory processes u-
Foucault, M. (1928). Les inhibitions internes de
Frankish, C. (1989). Perceptual organization and
Garner, W.G. (19-
Gardiner, J.M., Craik, F.I.M., & Birtwistle, J.
Gallistel, C.R. (1999). The organization of new-
Gallistel, C. (1990). The organization of learn-
Fischler, I., Rundus, D., & Atkinson, R.C.
Glanzer, M. (1972). Storage mechanisms in r-
Gigerenzer, G. (2000). Adaptive thinking: Rationa-
Fischler, I., Rundus, D., & Atkinson, R.C.
Frankish, C. (1989). Perceptual organization and
Garner, W.G. (19-
Gardiner, J.M., Craik, F.I.M., & Birtwistle, J.
Gallistel, C.R. & Gibbon, J. (2000). Time, rate,
eductive distributed model of ordering in s-
eductive distributed model of ordering in s-


lia, Nedlands, Australia.
Pollio, H.R., & Deitchman, R. (1964). The activation characteristics of a serial cognitive structure having oppositional end points. Mimeographed manuscript, University of Tennessee.


Watkins, M.J., Neath, I., & Sechler, E.S. (1989). Recency effect in recall of a word list when an immediate memory task is performed after each word presentation. American Journal of Psychology, 102, 265-270.


