MEMORY, 2001, *9* (4/5/6), 301–311

Modelling serial position curves with temporal distinctiveness

Jeffrey N. Rouder
University of Missouri-Columbia, USA

Pablo Gomez
Northwestern University, USA

We offer a critique of the temporal distinctiveness model of serial position effects (Nairne, Neath, Serra, & Byun, 1997). The temporal distinctiveness model combines a precise definition of stimulus distinctiveness with a memory perturbation process. The critique is empirically motivated—we show that with a more complete analysis, the temporal distinctiveness model does not adequately account for Nairne et al.'s experimental data. To better account for the data, we independently modified two components of Nairne et al.'s model: the mathematical form of the definition of temporal distinctiveness and the mathematical form of the mapping from distinctiveness to free-recall probabilities. Both of these modifications provided for better fits. Yet both Nairne et al.'s definition and our modified definition are fairly arbitrary. We show that a significant challenge to this approach is to find theoretically motivated constraints of the temporal distinctiveness model while providing for adequate fits to data.

When participants are asked to free-recall a list of previously studied items, they tend to produce more items from the beginning and end of the list than from the middle of the list. These findings, respectively termed primacy and recency, have traditionally been cited in support of Atkinson and Shiffrin's (1968) two-store memory model (see, for example, Rundus, 1971). The gist of the argument is that end-of-list items are recalled more frequently than middle-of-list items because the recall of end-of-list items is from both the short-term store and the long-term store, whereas recall of the middle-of-list items is from only the long-term store. Beginning-of-list items are better recalled than middle-of-list items because at the beginning of the list there is a smaller mnemonic burden and hence better transfer from the short-term store to the long-term store. Several alternative explanations of primacy and recency (e.g., Craik & Lockhart's, 1972, levels of processing framework, or Baddeley & Hitch's, 1977, working memory framework) are closely related to the two-store model in that explanations of primacy and recency are based on the assumption that different mnemonic processes operate at different serial positions. For example, in the levels of processing framework, recency effects occur because shallow traces are available for end-of-list items only. Neuropsychological research yields additional support for the notion that recency and primacy are supported by separate stores and/or separate processes for auditory stimuli. Patients with damage to anterior perisylvian region exhibit preserved recency but diminished primacy. But patients with damage to parts of the inferior parietal lobule exhibit the opposite result: pre-

Requests for reprints should be sent to Jeffrey N. Rouder, 210 McAlester Hall, Department of Psychology, Columbia, MO, USA, 65211.

This research was supported by National Science Foundation Grant BCS-9917561 to J. Rouder and National Institute of Mental Health Grants HD MH44640, National Institute for Deafness and other Communications Disorders Grant R01-DC01240, and National Science Foundation Grant SBR-9221940 to R. Ratcliff and G. McKoon. We are grateful to Ian Neath for providing detailed information on the temporal distinctiveness model.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210042000102

served primacy but diminished recency (Gordon, 1983). These results are consistent with the idea that primacy and recency may be mediated by different brain structures.

Nairne, Neath and their colleagues (Nairne et al., 1997; Neath, 1993) propose an alternative account of primacy and recency effects based on a single mnemonic process operating on a single store. The account is based on the observation that items at the beginning and end of the list are different from items in the middle of the list. Items at both ends of the list are presented near unique temporal events (namely, the beginning and end of the list). These items, therefore, are more *temporally distinct* than middle-of-list items. Participants encode the temporal position of an item as part of the general context and use the temporal distinctiveness of these beginning- and end-of-list items to aid in free-recall. The idea that distinctiveness is an important determinant in free-recall comes from interference theories of memory (e.g., Einstein, McDaniel, & Lackey, 1989; Postman & Stark, 1969). Interference from competing items explains omission errors in free-recall. Distinctive items are relatively immune to interference and are less likely to be omitted.

To account for variability in performance, Nairne, Neath, and colleagues assume that there is flux or change in memory through time (e.g., Estes, 1997). The record of an item's temporal position may change randomly after encoding. For example, at recall, the participant may mistakenly remember the first item as being in the second position. In this case, the temporal distinctiveness of the first item would be lessened. Beginning-of-list items are more likely to have their temporal positions confused (they need to be retained longer) than end-of-list items, and hence the temporal distinctiveness of beginning-of-list items is less than that of the end-of-list items. This explains why recency effects are larger than primacy effects for immediate free-recall tasks.

Temporal distinctiveness is not a new theoretical construct; it can be traced back to Murdock (1960). What is new is that Nairne et al. (1997, henceforth referred to as NNSB) have incorporated temporal distinctiveness in a rigorous and parsimonious formal model which accounts for primacy and recency effects under a number of experimental conditions. Because the model offers an elegant explanation of primacy and recency effects without postulating multiple mnemonic stores or processes, it merits further evaluation. In this paper, we demonstrate that the NNSB model-fitting endeavour is flawed and fails to capture several aspects of NNSB's own experimental data. We present two different ways of refining the NNSB model so that it does indeed provide a better fit to the data. In the process of refining the model, we demonstrate that some of the original NNSB assumptions, as well as our refinements, are arbitrary and lack theoretical backing. The unfortunate state of affairs is that the goodness-of-fit of the NNSB model is largely dependent on the particular forms of these atheoretical assumptions. Therefore, fitting the NNSB data with the NNSB model does not provide a satisfactory test of the core components of the temporal distinctiveness framework. Our conclusion is that temporal distinctiveness is not yet sufficiently developed to provide a satisfactory alternative to the modal two-store model.

THE TEMPORAL DISTINCTIVENESS DIFFUSION MODEL

It is easiest to understand the NNSB model within the context of a specific experimental paradigm. NNSB performed a number of free-recall experiments in which participants studied lists of letters and digits. Letters served as targets for a subsequent free-recall test while digits served as distractors. Participants were not required to commit the distractors to memory. There were always six target letters, but the number of distractor digits varied from condition to condition in two ways. First, the number of distractor digits between target letters was manipulated from two to twelve digits. Second, the number of distractor digits following the last target was also manipulated from two to twelve digits. NNSB did not manipulate these factors in a factorial manner; instead they chose nine different pairings. The conditions can be represented as ordered pairs in which the first argument denotes the number of distractors between targets, and the second argument denotes the number of distractors after the last target. The nine conditions were (2,2), (2,4), (2,8), (2,16), (2,24), (4,2), (8,2), (16,2), (24,2). The first five conditions are used to explore the effects of increasing the number of distractors after the presentation of targets, while the last four conditions are used to explore the effects of increasing the number of distractors between targets.

It is standard to plot the probability of an item being recalled as a function of its serial position during study. Such plots are referred to as serial position curves. A rising left tail indicates primacy effects whereas a rising right tail indicates recency effects. In NNSB Experiment 1, each list contained six targets, and hence there were six points for each serial position curve. Figure 1 shows these empirical serial position curves. Each of the nine panels corresponds to a different condition, and the small circles are correct recall proportions.

It is helpful to have confidence intervals around the free-recall proportions in evaluating formal models. Unfortunately, NNSB did not provide this information. Instead, we have "guesstimated" .95% confidence intervals in a manner described in the Appendix. These "guesstimates" are graphed as error bars in the figure. As discussed in the Appendix, the "guesstimates" are most likely larger than the true confidence intervals. By overestimating the confidence intervals, we are more confident that if several predictions are outside their respective confidence intervals, then the model is misspecified.

The NNSB model is predicated on three assumptions: first, that the temporal position of each item is stored; second, that the stored value of temporal position randomly changes in time; and third, that free-recall probabilities are based on the distinctiveness of these randomly-changed temporal positions. To implement the first assumption we assume that the participant accurately stores the temporal position of the item at the time of study. For example, when the fifth

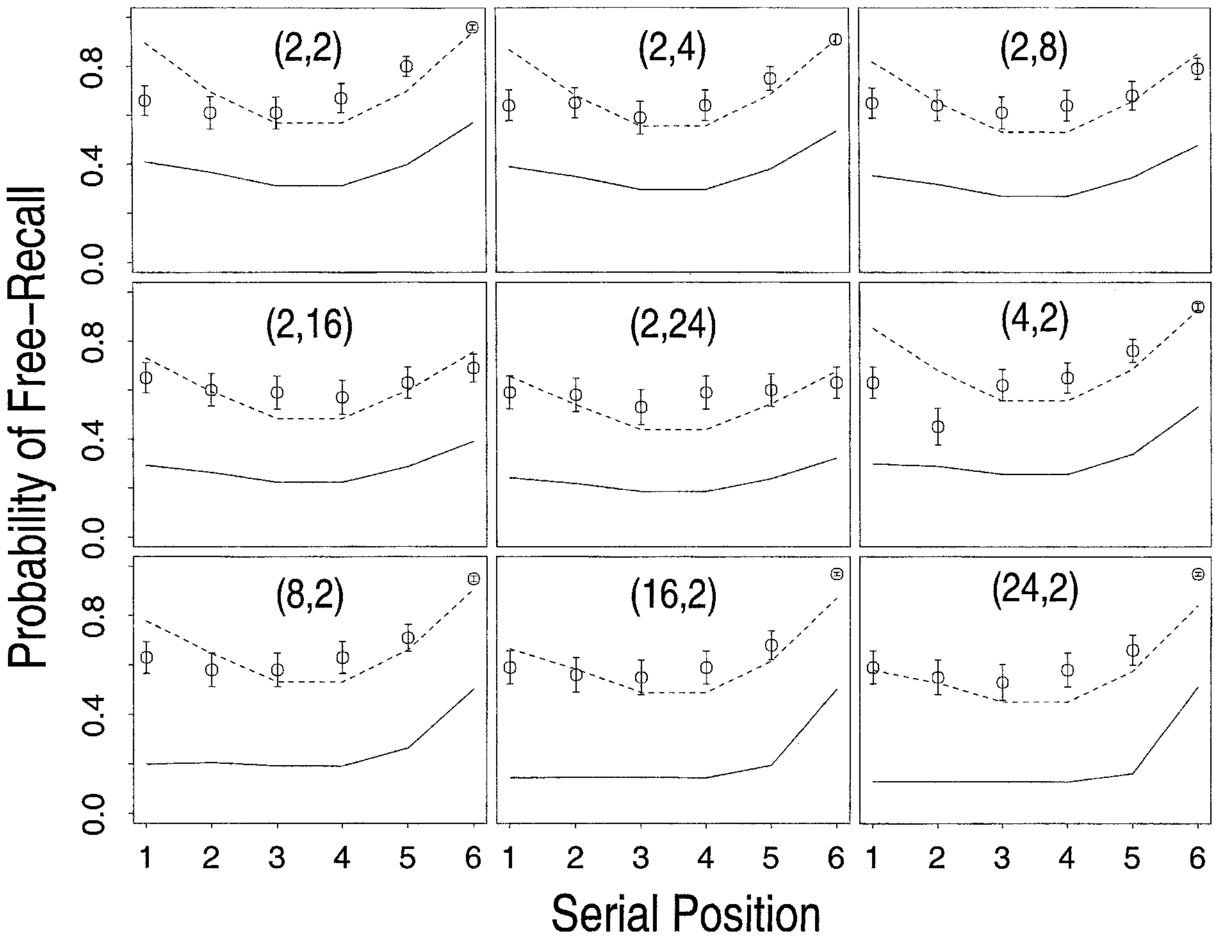

Figure 1. Serial position curves from NNSB's Experiment 1. The small circles are the data. The solid lines that grossly underestimate the data are from the temporal distinctiveness model evaluated with the parameters NNSB used to fit recency slopes. The dotted lines are fits from the temporal distinctiveness model with optimised parameters.

target is studied, the temporal position value of 5 is stored.[1]

The second assumption is that the temporal position of items can change. This is implemented by assuming that the temporal position of items undergoes a perturbation process or random walk (e.g., Estes, 1997). Suppose the stored temporal position of an item is x. A short time later, the stored temporal position may change to $x + 1$ with probability $\theta/2$, where θ is a free parameter. The stored temporal position may also change to $x - 1$ also with probability of $\theta/2$. Alternatively, the position may not change with probability $1 - \theta$. There are exceptions to those probabilities which occur when the temporal position is at the beginning of the list ($x = 1$) or at the end of the list ($x = 6$ for the NNSB experiments). In these cases, the temporal position remains unchanged with probability $1 - \theta/2$ or it can move one position towards the middle of the list with probability $\theta/2$. This change process is known as a random walk with reflecting boundaries (e.g., Feller, 1965). The number of changes in temporal position is directly related to the time between the study of an item and the start of recall. Therefore, end-of-list items undergo fewer changes on average than beginning-of-list items.

The most intriguing assumption is the third one, in which these perturbed temporal positions are used to calculate free-recall probabilities. The temporal distinctiveness of each item, denoted D_i, is the sum of distances between the item's temporal position and each of the other items' temporal positions.

$$D_i = \sum_{j=1}^{6} |x_i - x_j| \qquad (1)$$

where x_i denotes the temporal position of the i^{th} target at the time of recall. The last step is to specify a function that relates distinctiveness (D_i) to free-recall probabilities. NNSB did not want the success of their model to hinge on tertiary assumptions like the form of the relationship between distinctiveness and free-recall probability. Their solution was to use simplicity as a guide, and they chose to scale distinctiveness linearly. The probability of free-recalling the i^{th} target (denoted P_i) is D_i/k, where k is the scaling parameter.

In its minimalist form, the NNSB model has only three free parameters: the probability of a change (θ), the number of opportunities that a temporal position change may occur per unit time, and the scaling factor (k). In their implementation, NNSB used a fourth free parameter. They assumed that the number of change-opportunities per unit time in the retention interval was different from that during presentation of target items. This assumption is plausible as there may be different mnemonic loads while maintaining information in the retention interval than while storing new information in the study period. It is reasonable in this framework to expect the rate of temporal positions change to vary with mnemonic load.

In their paper NNSB did not fit the model to the serial position curves. Instead, they derived a quantity from their data called the *recency slope* (see Glenberg, Bradley, Kraus, & Renzaglia, 1983). The recency slope is the slope of the best fitting line through the correct free-recall proportions of the three most recent items (or the three right-most points in Figure 1). Although fitting the recency slopes is impressive, these fits do not imply that the model actually fits the serial position curves. The solid lines in Figure 1 show the model's predictions for the nine serial position curves for the parameters used by NNSB. These fits to the serial position curves are poor and grossly underestimate the correct free-recall proportions. To quantify the fit, we used a root-mean-squared (RMS) error statistic. The root-mean-squared error is the average discrepancy between the predicted probability of correct free-recall and the proportion of correct free-recall. When the parameters reported by NNSB are used to predict serial position curves the RMS error is .38, which is unsettling. It is evident from inspection that the predicted recency slopes fit the empirical recency slopes (this was the main empirical support NNSB provided for the model), but the gross underestimation of the serial position curves obviates this close fit.

[1] There are a few ambiguities in the NNSB article about whether distractors are items that receive temporal position assignments. In describing and illustrating the model, NNSB assigned temporal positions to both targets and distractors (see NNSB, p. 157). Hence the temporal position of the fifth target would be one more than the total number of preceding targets and distractors in the list. But, in fitting the temporal distinctiveness model to data, NNSB did not assign serial positions to distractors (Neath, personal communication). Therefore, as stated earlier, the fifth target was assigned temporal position 5. There is no a priori reason to prefer either approach. We follow the latter NNSB approach of assigning temporal position only to targets because this version of the model is sufficiently tractable that the parameters can be optimised to fit data.

The NNSB choice of parameters to evaluate their model was motivated by fitting the recency slopes. To find parameters that best fit the entire serial position curves, we used the SIMPLEX routine[2] (Nelder & Mead, 1965). The predictions from the best-fitting parameters are shown in Figure 1 as dotted lines. Although there is a capital improvement over the NNSB fit, the model still does a relatively poor job of accounting for the data (the root-mean-squared error was .099). Both NNSB's original parameters and the optimised parameters are shown in the first two rows of Table 1. These rows appear under the subheading "Murdock-NNSB Distinctiveness".

The bottom line is that the NNSB model fails to provide a satisfactory fit to the serial position curves. Even when the parameters are optimised, there are still large and systematic misses. The preceeding analysis shows how researchers may be led astray if they focus on derived statistics in the data at the expense of more fundamental ones. In the remainder of the paper, we attempt a few revisions of the NNSB model.

FROM DISTINCTIVENESS TO FREE-RECALL

Are there any models that keep the core assumptions of the framework and that adequately fit the data? The core assumptions are: (a) temporal position is encoded, (b) temporal position perturbs, and (c) the distinctiveness of the perturbed positions is predictive of free-recall performance. To answer this question, we generalised the tertiary assumption about the form of the function that relates distinctiveness to free-recall while leaving intact the core assumptions. In the NNSB model, distinctiveness is scaled by a constant (k), e.g., $P_i = D_i/k$. This function relating distinctiveness to recall probability can be generalised to an affine form, e.g., $P_i = k_0 + D_i/k$. The model with this function relating probability to distinctiveness was fit (see the dotted lines in Figure 2), and the resulting root-mean-squared error was 0.044. The best fitting parameters for this model are presented in Table 1 in the row labelled *Affine scaled*. The model fits much better with the affine scaling function than with the original linear one.

Establishing that there exists some scaling of distinctiveness to free-recall probability that, when coupled with the core assumptions, provides for a good fit of the data is a critical first step. If we were unable to provide such an existence demonstration, we could have rejected the core of the framework. But this existence result is only a first step. There is a serious problem in leaving the functional form unconstrained in theory. If researchers are unconstrained and find the best functional form to fit the data, one is never sure that the core components can be rejected. For any given set of data, there may exist some function form of the scaling from distinctiveness to free-recall probability that, when coupled with the core assumptions, will provide an adequate fit. Fitting

TABLE 1
Parameters and fits for NNSB models

	θ	*Change-opportunities per second* In-between targets	*Recency*	*Scale* k	k_0	*Threshold*	γ	*RMS error*
Murdock-NNSB Distinctiveness								
Original parameters	.05	17	7	22	—	—	—	.3761
Optimised parameters	.027	3	7	15.18	—	—	—	.0984
Affine scaled	.056	56	12	23.30	.492	—	—	.0438
Response process	.059	11	5	—	—	8.66	—	.0878
Flexible Distinctiveness								
Scaled	.149	5	3	5.49	—	—	.85	.0428
Response process	.223	5	3	—	—	9.107	1.56	.0691

[2] The NNSB model is ill-suited to optimisation in SIMPLEX because two of the parameters, the number of change-opportunities per unit time in the recency interval and the interval between targets, are integer valued. We used the following approach to mitigate this problem. First, a grid was set up over a large range of numbers for the change-opportunity parameters. For each point on the grid, we estimated the other two parameters with SIMPLEX and calculated the RMS error. The point on the grid with the lowest RMS then served as the best values for the two change-opportunity parameters.

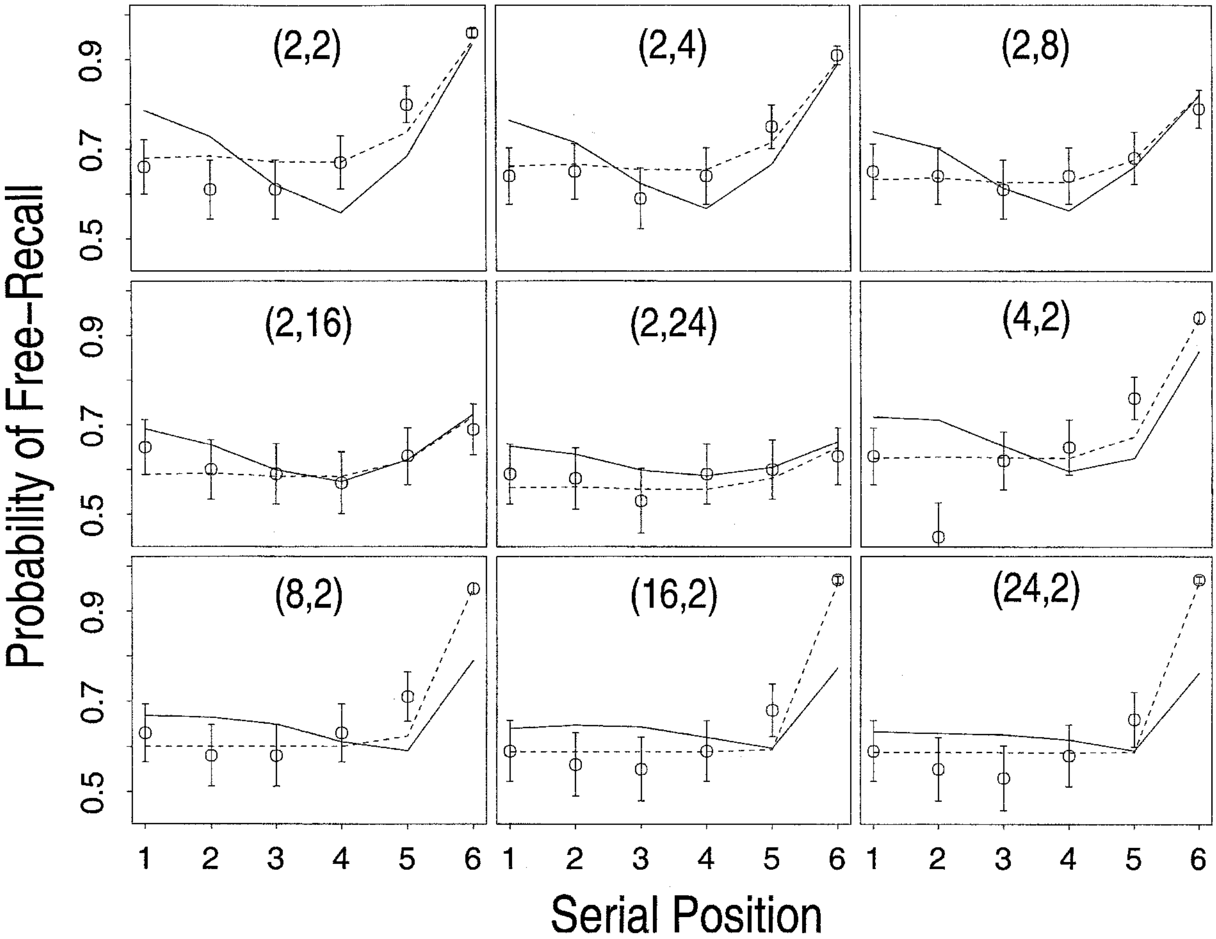

Figure 2. Fits of the affine scaled NNSB model (dotted lines) and the NNSB model with a theoretically grounded response process (solid lines).

the model with an unconstrained functional form may be more an exercise in curve fitting than a test of psychological theory.

Both the original linear scaling function and our affine scaling function are arbitrary in nature. There is no reason to prefer one over the other. The next step in testing the core assumptions is to postulate a theoretically grounded response process for relating distinctiveness to free-recall probabilities. We implemented a threshold approach. If, on a given trial, the distinctiveness of an item was greater than a threshold, it was sufficiently distinct to be immune to interference from other items and was recalled. Conversely, if the temporal distinctiveness was below the threshold, then the participant failed to recall that item due to interference. The mechanics of this response process are very similar to the Theory of Signal Detection and are used in other memory models such as SAM (Gillund & Shiffrin, 1984), TODAM (Murdock, 1982, 1983), and Atkinson and Juola's (1973, 1974) two-process model. It would be a strong demonstration if, when coupled to a theoretically grounded response process, the core assumptions of the temporal distinctiveness framework can account for the data.

To understand how this response process operates, it is necessary to examine how NNSB calculated distinctiveness values. At the time of test, an item has undergone a number of temporal position perturbations. In fact, because the perturbations are random, the temporal position of an item at test is represented by a discrete probability distribution with mass at several temporal positions. NNSB used the mean temporal position to calculate average distinctiveness. This approach of computing an average distinctiveness forces the researcher to specify a function to relate average distinctiveness to free-recall response probability. Of course, the choice of the form of the function is arbitrary. But, as has been shown, the form of this function greatly influences the fit of the model.

The major advantage to the threshold-based alternative is that there is no need to specify a functional form to relate average distinctiveness to response probability. To calculate free-recall probabilities, one must calculate the whole distribution of distinctiveness from the whole distribution of temporal positions. Free-recall probabilities correspond to the probability mass of distinctiveness that is greater than a criterion. Monte Carlo simulation was used to calculate the distribution of temporal positions. Based on this simulated distribution, the probability mass of distinctiveness was calculated as well as the proportion that exceeded the criterion. This approach is far more principled than the NNSB approach because the trial-by-trial variability in the perturbation process drives the variability in the distinctiveness distribution and, consequently, the proportion of that distribution over threshold.

We used SIMPLEX in conjunction with Monte Carlo simulations to find the best-fitting parameters. The results are shown as solid lines in Figure 2, and the best-fitting parameters are shown in Table 1 in the row labelled *Response process*. The fit is not much better than the original NNSB model with optimised parameters. The conclusion is that when coupled with a theoretically grounded method of mapping distinctiveness to response probability, the core assumptions provide at best a rough fit to the data. Certainly, there are systematic deviations, and we do not find the fit sufficient.

FROM SERIAL POSITION TO TEMPORAL DISTINCTIVENESS

In this section, we search for a means of revising one of the core assumptions. Up to this point, the functional form of temporal distinctiveness was considered part of the core assumptions. Temporal distinctiveness in the NNSB and Murdock formulations is the sum of position differences (see Equation 1). The intuition captured in this formulation is that (a) the beginning and end of the list are more distinct than the middle, and (b) distinctiveness is symmetric, favouring the beginning and end equally. One advantage of this formulation is its simplicity—it requires no parameters. The drawback is that the function is rather rigid. The approach pursued here is to find a temporal distinctiveness function that preserves the symmetry and bowed shape of the sums-of-differences function, but is more flexible. One such function, henceforth termed the *flexible distinctiveness function*, is:

$$D_i = \left[\frac{x}{n+1}\left(1-\frac{x}{n+1}\right)\right]^{-\gamma} \quad \gamma \geq 0. \tag{2}$$

In the flexible distinctiveness function, x is the temporal position of an item at recall and n is the list length. Both the Murdock-NNSB temporal distinctiveness function and the new flexible distinctiveness function are shown in Figure 3. The reason for the $(n + 1)$ term in the denominators in Equation 2 is to preserve symmetry. The exponent of the power function (γ) acts as a tuning parameter. Large values correspond to a sharp tuning in which both beginning- and end-of-list items are much more distinct than middle-of-list items. Small values correspond to a shallow tuning in which beginning- and end-of-list items are only somewhat more distinct than middle-of-list items.

It is important to recognise a lack of theoretical grounding in flexible distinctiveness. A key concept in the Murdock-NNSB distinctiveness is that distinctiveness of an item is a function of the temporal position of the item as well as the temporal positions of all other items in the list. Brown, Neath, and Chater (1998) termed the Murdock-NNSB distinctiveness as *global* because of this property. They critiqued[3] global properties in distinctiveness formulations and recommend a *local* distinctiveness in which the distinctiveness is a function of the item's temporal position and its nearest neighbours.

The flexible distinctiveness formulation is neither global nor local; the distinctiveness of an item depends solely on its temporal position without regard to the positions of other items. The best way of viewing flexible distinctiveness is as a form that may mimic other formulations of local or

[3] Brown et al. (1998) presented participants with serial lists in which the time between items was varied within the list. There was a large temporal gap between the middle item and all of the others. The Murdock-NNSB distinctiveness predicts greater primacy and recency in this paradigm than in the one in which the time between items is constant. In Brown et al.'s experiments, there was an increase for correct recall for the middle item at the expense of the primacy and recency. Brown et al. concluded that distinctiveness was local in nature, and because the isolate in the middle of the list had no near neighbours, it was distinct. But, for the paradigm presented here, there are no differential predictions between global and local temporal distinctiveness. Both global and local distinctiveness predict symmetric u-shaped functions.

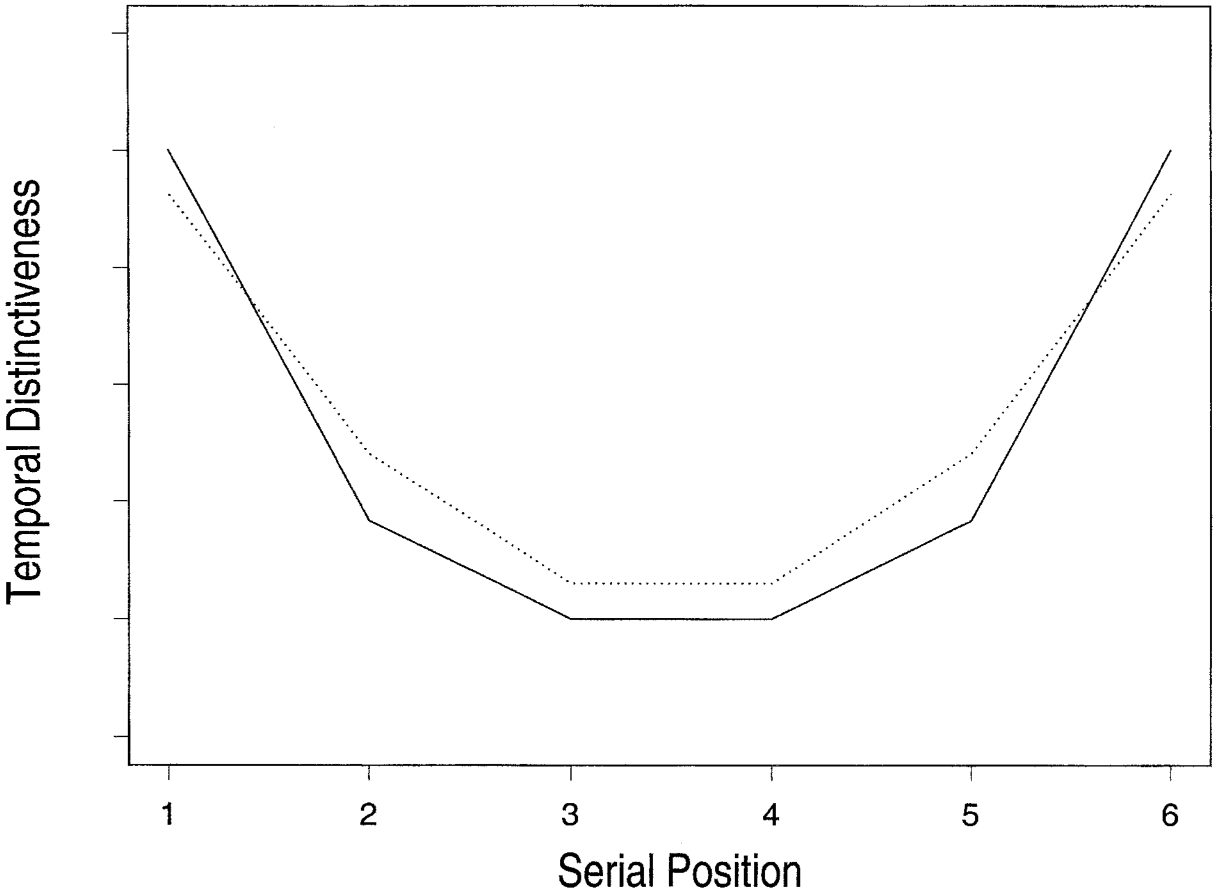

Figure 3. Temporal distinctiveness functions. The solid line corresponds to flexible distinctiveness (Equation 2) evaluated with γ = .85, and the dotted line corresponds to the Murdock-NNSB temporal distinctiveness function (Equation 1).

global temporal distinctiveness for this task. Flexible distinctiveness is designed to mimic u-shaped, symmetrical distinctiveness functions. Therefore, if a model with flexible distinctiveness does not account for the data, we can be fairly confident that this failure is not due to a mis-specification of the u-shaped temporal distinctiveness function. Instead, a failure represents a deeper failure of the core assumptions of the perturbation-distinctiveness approach.

The first step is to provide an "existence" demonstration. We show that there exists a functional form from distinctiveness to free-recall probabilities that, when coupled with the core assumptions and flexible distinctiveness, can fit the data. We used a linear function to relate distinctiveness to free-recall probabilities. The resulting model has five parameters, θ, the two change opportunities per unit time parameters, k (the multiplicative scaling parameter), and γ (the tuning parameter). As shown by the dotted lines in Figure 4, this model provides a relatively satisfactory account of the NNSB serial position curves (rms = .043). The best fitting parameters are shown in Table 1 under the sub-heading *Flexible distinctiveness* in the row labelled *Scaled.*

As mentioned before, these fits are necessary, but not sufficient. The good fit may be a function of the arbitrary relationship between distinctiveness and free-recall. As a final test, we coupled flexible temporal distinctiveness with the more theoretically motivated response process discussed earlier. The best-fitting parameters are shown in Table 1 (in the row labelled *Response process*), and the predictions from these parameters are shown as solid lines in Figure 4. Although this fit is better than the corresponding fit with Murdock-NNSB distinctiveness (this is not surprising as there is an additional parameter), the fit still misses some points (rms = .070).

This last model fit (represented by the solid lines) is the fairest test of the temporal distinctiveness framework. We have generalised temporal distinctiveness while still requiring the u-

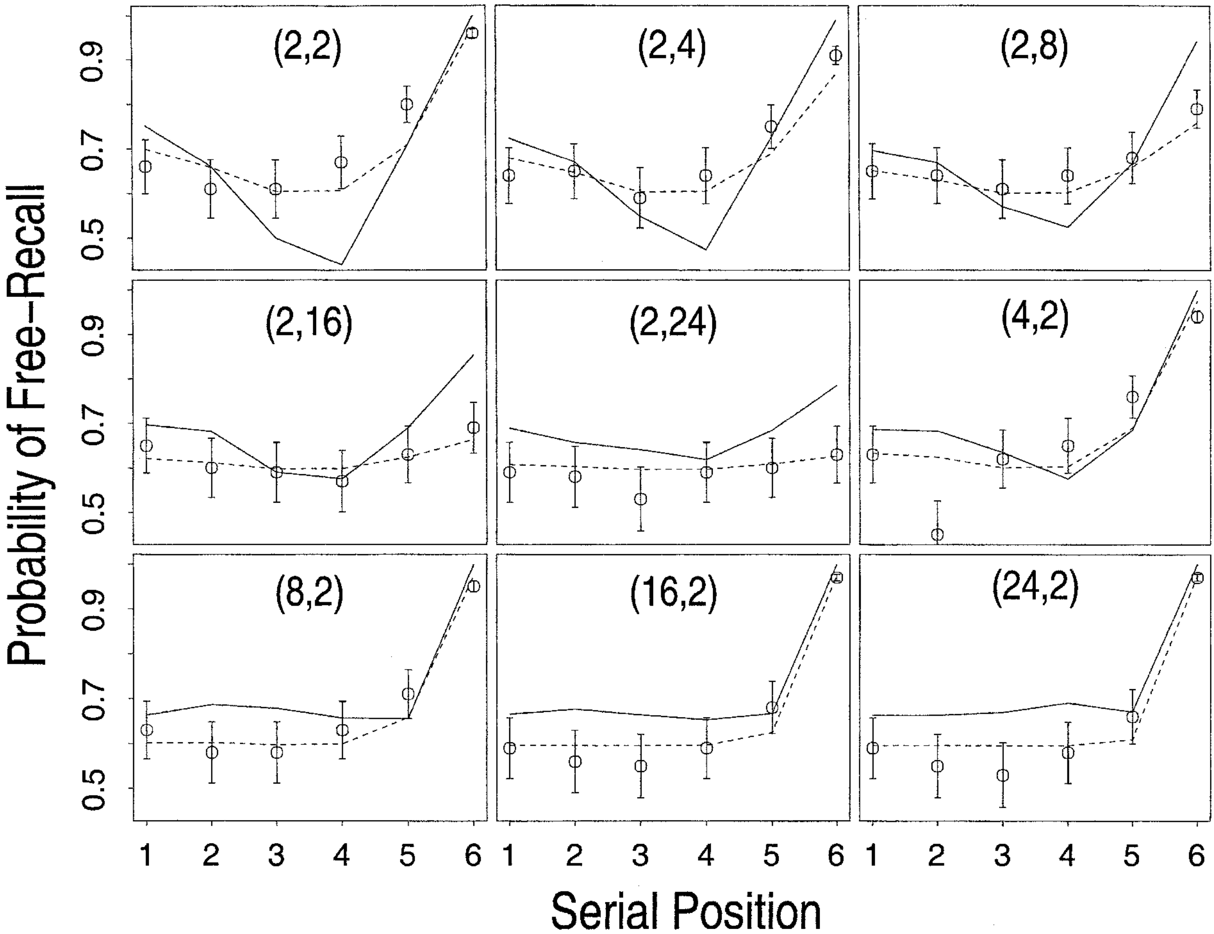

Figure 4. Fits to models with flexible distinctiveness (Equation 2). The dotted lines are for the flexible distinctiveness model with the original NNSB scaling. The solid lines are for the flexible distinctiveness model with the theoretically motivated response process.

shaped and symmetry properties. We have have implemented a theoretically grounded response process such that the fit is not a function of the flexibility of tertiary assumptions. With only a "guesstimate" of the variability in the data, evaluative statements must be made cautiously. On one hand, this is a good fit for a field with a history of accounting for only summary trends in data. On the other hand, the systematic nature of the misses leaves us unsatisfied. In the end, we suspect that those inclined to favour the model may view our fits as somewhat supportive of the temporal distinctiveness framework. Those less inclined to favour the framework may be less convinced that the fits are acceptable.

SUMMARY

The conclusion of the preceding analyses are threefold. First, the NNSB temporal distinctiveness model fails to account for serial position curves. Second, there are ways of modifying the NNSB temporal distinctiveness model that yield much better fits. We were able to generalise the scaling from distinctiveness to free-recall performance. Although this generalisation fits better, it is quite arbitrary and points to perhaps the most important issue—the fit of the model depends on tertiary and arbitrary assumptions. When we coupled the core assumptions of the temporal distinctiveness framework to a more theoretically grounded response process, we were unable to provide satisfactory fits. Third, the same basic pattern of results holds when the form of temporal distinctiveness is generalised. That is, the core framework provided satisfactory fits with arbitrary assumptions but failed when coupled with a theoretically grounded response process. In sum, temporal distinctiveness models may have the potential of providing a single process account of serial position effects that is elegant and parsimonious. Yet, as shown here, they have not been shown to fit benchmark data without recourse to arbitrary, and probably all-too-flexible, assumptions.

REFERENCES

Atkinson, R.C., & Juola, J.F. (1973). Factors influencing the speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and Performance IV* (pp. 583–612). New York: Academic Press.

Atkinson, R.C., & Juola, J.F. (1974). Search and decision processes in recognition memory. In D.H. Krantz, R.C. Atkinson, R.D. Luce, & P. Suppes (Eds.), *Contemporary development in mathematical psychology (Vol 1)* (pp. 242–293). San Francisco: Freeman.

Atkinson, R.C., & Shiffrin, R.N. (1968). Human memory: A proposed system system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory. Vol. 2* (pp. 89–195). New York: Academic Press.

Baddeley, A.D., & Hitch, G.J. (1977). Recency revisited. In S. Dornic (Ed.), *Attention and Performance 6*, (pp. 647–667). Hillsdale, NJ: Lawrence Earlbaum Associates Inc.

Brown, G.D.A., Neath, I., & Chater, N. (1998). *Scale-independent memory and perception*. Paper presented at the 39th Annual Meeting of the Psychonomic Society. Dallas, TX.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Experimental Psychology: General*, *104*, 671–684.

Einstein, G.O., McDaniel, M.A., & Lackey, S. (1989). Bizarre imagery, interference, and distinctiveness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 137–146.

Feller, W. (1968). *Introduction to probability theory and its applications* (Vol. 1, 3rd ed.). New York: Wiley.

Estes, W. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review*, *104*, 148–169.

Gillund, G., & Shiffrin, R. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 97–123.

Glenberg, A.M., Bradley, M.M., Kraus, T.A., & Renzaglia, G.J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*, 231–255.

Gordon, W. (1983). Memory disorders in aphasia: I. Auditory immediate recall. *Neuropsychologia*, *21*, 325–339.

Johnson, N.L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions. Volume 2 (Second ed.)*. Chichester, UK: Wiley.

Murdock, B.B. (1960). The distinctiveness of stimuli. *Psychological Review*, *67*, 16–31.

Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609–626.

Murdock, B.B. (1983). A distributed memory model for serial-order information. *Psychological Review*, *90*, 316–338.

Nairne, J.S., Neath, I., Serra, M., & Byun, E. (1997). Positional distinctiveness and the ratio rule in free recall. *Journal of Memory and Language*, *37*, 155–166.

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

Nelder, J.A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.

Postman, L., & Stark, K. (1969). Role of response availibility in transfer and interference. *Journal of Experimental Psychology*, *79*, 168–177.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 63–77.

APPENDIX

NNSB published only mean correct free-recall proportions for their Experiment 1. We provide a reasonable "guesstimate" of confidence intervals. Our approach is to assume that each participant's recall proportion is a sample from a beta distribution (Johnson, Kotz, & Balakrishnan, 1994). The beta distribution is a functional form that can describe densities with high variability on the [0, 1] interval. The beta is described by two parameters α and β. We set $\beta = 1.2$ for all conditions and serial presentation positions. We then adjusted α such that the expectation of the beta density equalled the observed response proportion. For example, the value of the response proportion for the first serial position in the condition (2, 2) is .66. By choosing $\alpha = 2.33$, the resulting beta density has an expected value of .66, matching the observed response proportion. Figure 5 shows such a density and, as can be seen, it is quite variable. In fact, it is far more variable than the differences in individual participants. The variance of this density is used to compute confidence intervals. Because the variance of the beta is an overestimation of variability across participants, the confidence intervals are inflated. This, in turn, makes it harder to reject ill-fitting models. If a model can be rejected by an ill fit, then the rejection is made with more confidence than would be if the confidence intervals were estimated without such bias.

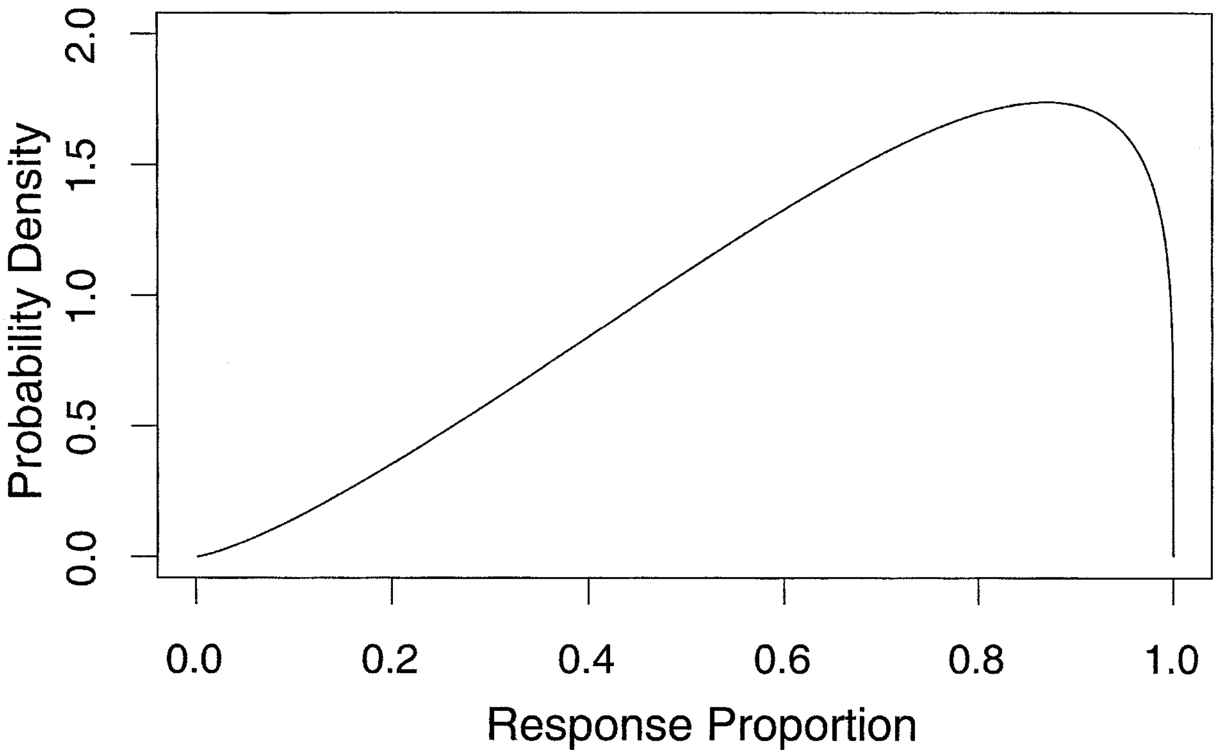

Figure 5. A beta-density model of variability across individuals in free-recall proportions. The figure shows the density for a mean response proportion of .66.