

A constrained Rasch model of trace redintegration in serial recall

STEVEN ROODENRYS AND LEONIE M. MILLER

University of Wollongong, Wollongong, New South Wales, Australia

The notion that verbal short-term memory tasks, such as serial recall, make use of information in long-term as well as in short-term memory is instantiated in many models of these tasks. Such models incorporate a process in which degraded traces retrieved from a short-term store are reconstructed, or redintegrated (Schweickert, 1993), through the use of information in long-term memory. This article presents a conceptual and mathematical model of this process based on a class of item-response theory models. It is demonstrated that this model provides a better fit to three sets of data than does the multinomial processing tree model of redintegration (Schweickert, 1993) and that a number of conceptual accounts of serial recall can be related to the parameters of the model.

The serial recall task, in which participants attempt to recall a series of sequentially presented items in the order in which they were presented, has been the most commonly used task to investigate verbal short-term memory. Many models of short-term memory, or simply of the serial recall task, incorporate a process by which degraded temporary memory traces of the items (words, digits, letters, or nonsense words) are reconstructed, or "redintegrated" (Schweickert, 1993), by accessing traces in long-term memory (see, e.g., Brown & Hulme, 1995; Nairne, 1990; Page & Norris, 1998b). Whether the initial degradation of the trace occurs through passive decay (see, e.g., Baddeley, 1986; Page & Norris, 1998b) or active interference (see, e.g., Nairne, 1990) is irrelevant to the redintegration process, but the existence of such a process seems likely, given experimental findings showing an influence of long-term memory factors on serial recall performance. For example, words, by definition, are distinguishable from nonwords in having lexical representations in long-term memory, and serial recall performance is better for lists of words than for lists of nonwords when they are matched on other factors (see, e.g., Hulme, Maughan, & Brown, 1991). High-frequency words differ from low-frequency words in terms of how often they have been encountered in the past, so any superior memory performance for high-frequency versus low-frequency words presumably reflects the long-term memorial impact of those encounters. Robust frequency effects are reported for the serial recall task when pure lists are used (see, e.g., Gregg, Freedman, & Smith, 1989; Hulme et al., 1997; Tehan & Humphreys, 1988; but see Hulme, Stuart, Brown, & Morin, 2003, and Morin, Poirier, Fortin, & Hulme, 2006, for discussions regarding effects in mixed lists). The aims of the present article are to describe a new measurement model of the redintegration process and compare it with an existing model, the multinomial processing tree model of Schweickert (1993).

The Multinomial Processing Tree Model of Redintegration

The multinomial processing tree (MPT) model posits a two-stage process (Schweickert, 1993). In the first stage, the trace is retrieved from short-term memory with a particular probability, i , of being intact. If the item is intact, it is recalled correctly; if it is not intact, an attempt is made to redintegrate the item with probability r of successfully reconstructing the correct item. This means that the probability of correctly recalling item x becomes the sum of the probability of that item's being intact and the product of the probability that it is not intact and the probability that the item is successfully reconstructed, as shown in Equation 1:

$$p(c_x) = i_x + (1 - i_x)r_x. \quad (1)$$

This model has been applied successfully to data showing a word frequency effect (Hulme et al., 1997) and to the effects of irrelevant speech on serial recall (Buchner & Erdfelder, 2005) by assuming that i , the likelihood of the short-term trace's being intact, decreases through the list, and that r , the likelihood of correctly redintegrating an item, varies across stimulus sets or conditions. For example, high-frequency words have a larger r than do low-frequency words, given the assumption that they are more accessible within long-term memory.

Although the MPT model has been shown to fit several sets of data, thus providing a good measurement model of redintegration, it includes a number of theoretical assumptions that might be questioned. The model treats the integrity of the short-term trace in a discrete fashion; that is, the short-term trace is either intact or it is not, and if the trace is not intact, it has the same likelihood of being redintegrated regardless of how degraded it is. In fact, the model does not incorporate any notion of *degree* of degradation. As a result, the model seems at odds not only

S. Roodenrys, steven_roodenrys@uow.edu.au

with common-sense notions of decay and interference, but also with those notions in more formal models (e.g., Nairne, 1990) in which degradation is at least more fine-grained and perhaps continuous, and with such notions as the gradual decline of activation.

A second issue with the multinomial model is that it does not specify the process that determines whether the short-term trace is intact or not. One possibility is that the traces maintain information about the length of the item, perhaps by using a positional phoneme coding. This would allow the multinomial model to deal with the situation in which the word *scat* is presented, but only *cat* is retrieved from the short-term store, if the retrieved representation also coded the missing phoneme position. However, it is still necessary to evaluate whether all of the phoneme positions are filled, and, as yet, there has been no discussion of how this might be done.

One further issue of concern with the MPT model is that it can never explain serial recall performance that falls to zero for more than one position. Of course, this does not happen in reported data very often, since experiments are usually designed to avoid such floor effects, but anyone who has tested participants using longer lists has encountered individuals who cannot recall any items for more than one position. The problem for the MPT model is that the minimal performance level in the model is equal to the redintegration parameter for a set of stimuli. Once there is no likelihood of a trace's being intact, performance equals the value of r . This means that zero recall might occur for nonwords (but see Thorn & Frankish, 2005, for evidence that even nonwords may be redintegrated), but it should not happen for a set of words; so, if list length were increased, performance on items in later positions would drop only as far as this nonzero minimum.

Thorn, Gathercole, and Frankish (2005) discussed a revision of the original MPT model of redintegration first described by Gathercole, Frankish, Pickering, and Peaker (1999). In this revision, they suggested that a trace can be entirely lost and that such a loss forms a third possibility for the initial state of the trace (i.e., intact and thus recalled correctly, lost and thus not recalled, and partially degraded and thus redintegration is attempted). Thorn et al. discussed the revised model with the aim of accounting for the actual proportion of errors that preserve some elements of the original stimulus, since these errors must be a product of the redintegration process. Adding a parameter for the probability that a trace is lost altogether allowed Thorn et al. to examine these types of errors in more detail, and it also addressed one of the issues raised above. This revision of the model does allow for performance to fall below the level of the r parameter, since the probability of recalling an item when there is no possibility of its being intact is $(1 - i - l) * r$, where i is the probability of the trace's being intact (i.e., 0, in this instance) and l is the probability of its being lost altogether. If l is nonzero, then the performance level will be below r .

The Thorn et al. (2005) revision of the MPT model does not address the concern with the discrete nature of degradation in the original model. The revised model allows for three states—intact, degraded, or lost—but this does

not equate to the notion of degradation as a continuous variable. Allowing for three states also creates an additional concern: that this version of the model is, in fact, too powerful. It adds another parameter, which allows the model to produce a perfect fit to the data, and Thorn et al. demonstrated that the model did produce a perfect fit to their data. In fact, because the model now has more free parameters than data points for any set of data, it is consequently capable of fitting any data. This can be seen as a weakness, in that it undermines the theoretical usefulness of the model (Myung & Pitt, 1997).

Modeling Trace Degradation As a Continuous Variable

Roodenrys (in press) has recently suggested an account of redintegration in which degradation is a continuous process, and all items retrieved from the short-term store undergo redintegration. Roodenrys argued that if information loss is gradual, then traces of varying integrity are retrieved from the short-term store. At the same time, items vary in how readily they can be redintegrated as a result of their properties in long-term memory as compared with the properties of other items, and this variability could be seen as a property of the item itself, as is assumed in the MPT model. Given these two assumptions, the effect of a change in degradation on the likelihood of recall may not be linear but rather greater for moderate levels of degradation, and the variation in the effect is not the same for all items. To explain, the effect of a small amount of degradation on an item that is more or less intact will have less impact than the same amount of degradation on an item that is already moderately degraded; for example, losing one additional phoneme from a seven-phoneme word would have less effect on recall if this left six phonemes of the word intact than if it left three phonemes intact. At the same time, if a trace is already severely degraded, further degradation should have minimal effect on the likelihood of recall.

This conceptualization of degradation describes a sigmoidal or ogive-shaped function relating trace integrity to recall performance, whose slope, which represents the decline in performance, is largest at more moderate levels of trace integrity/performance. Such an intactness profile operates for all types of items; however, the redintegrability of items will influence the degree to which overall performance holds up in the face of degradation. For example, items that are more easily redintegrated, such as high-frequency words, would follow the same degradation pattern as low-frequency words, but on a curve shifted to the right on an axis representing trace degradation, with redintegrability counteracting to some extent the actual degradation of items up to the point at which degradation is complete and there is effectively no short-term trace left.

We propose a model of redintegration in which both short- and long-term memory provide information about the identity of the to-be-recalled item, regardless of the degree of integrity of the short-term trace. We suggest that the trace retrieved from the short-term store is used to activate information in phonological long-term memory and that the information retrieved from long-term memory is

used to produce a response. In this model, the state of the information in long-term memory does not vary across serial position, whereas the information from short-term memory does vary according to serial position. This is much the same as in the MPT model, but in this case, the short-term factor represents the integrity of the trace rather than the likelihood of its being intact. In this way, the integrity of the trace is treated as a situation-specific factor, whereas the ease-of-reintegration factor is treated as being item specific, applying in all positions for that item. In the model, the probability of correctly recalling item x , $p(c_x)$, is given by the equation

$$p(c_x) = \frac{i_x r_x}{i_x r_x + (1 - i_x)(1 - r_x)}, \quad (2)$$

where i is the integrity or intactness of the short-term trace at the time of recall, and r is how readily reintegrated the item is. It should be noted that these variables are not probabilities, as in the MPT model, but rather are properties of the representation of the item in the short- and long-term memory systems, respectively. Thus, for a particular item in an experiment, i will vary across serial positions in a list as a function of interference or decay, but r will not vary.

Interestingly, this equation, in different guises, has a diverse history in psychology. In the form presented above, it is identical to the fuzzy logic model of perception of Massaro and colleagues (see, e.g., Massaro, 1987; Massaro & Cohen, 1983). In that model, the two variables relate to different sources of information about the identity of a phoneme, such as visual and auditory sources. However, Crowther, Batchelder, and Hu (1995) demonstrated that the fuzzy logic model is also equivalent to one member of a class of item-response theory models from psychometrics known as Rasch models, originated by Rasch (1960). In that model, the two parameters represent subject ability and item difficulty. In the model we propose, the two parameters represent the integrity of the short-term trace and a long-term memory factor analogous to difficulty of reintegration. Furthermore, Crowther et al. showed that the Rasch models can be reduced to a simple logit model. The function underlying these models has the sigmoidal shape and the properties described above. For the present purposes, however, we believe it is preferable to use the equation above, since it limits the parameters to values between 0 and 1; we will refer to the model as the constrained Rasch (CR) model. Doing this makes comparison with the MPT model slightly easier and maintains a model form that is consistent with the concept of intactness that ranges from perfect to absolute degradation.

Comparing the MPT and CR Models

Before we report the fit of the two models to some data sets, there is an issue with these models that warrants some discussion. Both the MPT model and the CR model are nonidentifiable. That is, for any given data set, there is not a unique set of parameters that provides the best fit to the data. Rather, an infinite number of parameter value sets provide an equally good maximal fit. Buchner and

Erdfelder (2005) pointed this out for the MPT model of reintegration, and Crowther et al. (1995) discussed the issue at some length in relation to the fuzzy logic model of perception. As Buchner and Erdfelder explained in the case of the MPT model, *nonidentifiability* means that the parameter estimates cannot be taken as actual probabilities for any set of i values, since a different set that fits just as well can be found by changing the r parameter appropriately. However, the ratio of differences on a parameter (either i or r , in this case) is uniquely determined by the data, making discussion of the relative values across levels of each parameter meaningful. For example, in the MPT, if i decreases across serial positions in the list, it is not possible to interpret the values as the actual probability of the short-term trace's being intact, since these values are not a unique best-fitting set. However, it is meaningful to talk about the difference in the i parameter across different positions, since the proportional change across positions is equivalent for all the possible solutions.

As Crowther et al. (1995) discussed, nonidentifiability creates a number of problems for interpretation of the parameter values in a model. As just described, the parameter set cannot be interpreted as the set that provides the best fit of the model to the data, and thus does not represent the true psychological values. In addition, in the case of experiments in which each of the two parameters represents a variable that has been factorially manipulated, it is not possible to evaluate the relative importance of the factors in determining performance without further mathematical transformation (see Crowther et al., 1995).

One solution to the nonidentifiability problem described by Crowther et al. (1995) and adopted by Buchner and Erdfelder (2005) is to arbitrarily fix one value of a parameter in the model, such as the r value for one experimental condition or set of stimuli. Although this does not make interpretation of the parameters any more meaningful, it does provide a unique set of parameters and allows us to compare the goodness-of-fit of the two models of interest in this article. Accordingly, this approach is taken in all of the following estimations. In the case of the CR model, this is done by setting the r parameter to .5 for one of the stimulus sets. In the case of the MPT model, all three data sets reported below have a minimum level of performance between .20 and .22. Setting the r parameter for the condition with the lowest recall level to .1 should allow the algorithm to converge on values that provide a stable solution.

Modeling Long-Term Memory Effects

We began our comparison of the two models by examining their fit to the data from Experiment 2 of Hulme et al. (1997), since that study showed that the MPT model provides an extremely good fit. If the CR model were unable to provide a similarly good fit, one would have to question its usefulness. All modeling was performed using the free-ware statistical programming language R. Hulme et al.'s (1997) Experiment 2 involved the auditory presentation and spoken recall of seven-item lists whose items were drawn either from a set of high-frequency words or from a

set of low-frequency words. The fit was achieved by setting the level of the r parameter for the low-frequency condition and allowing the r parameter for the other condition to vary and by varying the i parameter across serial positions but using the same values across the different conditions.

The parameter values and fit statistics for both the MPT model and the CR model are shown in Table 1, and the fit against the data is shown in Figure 1. Both models contain eight free parameters to fit 14 data points, and as Table 1 and Figure 1 show, both models provide a very good fit to the data. The r^2 values for the two models are extremely high and almost identical. The G^2 goodness-of-fit statistic is also almost identical for the two models.

When comparing models with different numbers of free parameters, it is possible to evaluate the significance of the difference in fit provided by the models by testing the difference in G^2 between the models (see, e.g., Buchner & Erdfelder, 2005). In the present case, this is not possible, since the two models have the same number of parameters, and consequently there is no difference in the degree of freedom for the models. Instead, it is possible to evaluate model fit by calculating the Bayesian information criterion (BIC) for each model, shown in Table 1, and then transforming these values to BIC model weights ($wBIC$), which can be interpreted as the probability that the model is the best model out of those compared (Wagenmakers & Farrell, 2004). For the MPT model, the $wBIC$ is .399; for the CR model, .601. This means that the CR model is approximately 1.5 times more likely to be the better model despite the small difference in the fit statistics between the models.

In the data shown in Figure 1, it is possible that a number of factors underlying performance may all be captured by the r parameters. For example, Roodenrys and Quinlan (2000) showed that recall was better when the same small set of words was used in an experiment, as compared with when an open pool of words was used without repetition. This suggests that repeated presentation of words across different lists makes them more available to the redintegration process; this would be captured in these models as a change in the r parameter. However, this effect interacted with word frequency: There was an effect of repetition for low-frequency word sets but not for high-frequency word sets. In addition, Stuart and Hulme (2000) argued that the frequency effect in serial recall is due to greater interitem associations between high-frequency words than between low-frequency words. We do not think that these other factors are solely responsible for the difference in the parameter values, but they may be contributing to it. It is possible to test the models against a set of data obtained from a different methodology in which the role of these factors, as well as some other potential confounds, is minimized.

A second set of data comes from the regression study of Roodenrys, Hulme, Lethbridge, Hinton, and Nimmo (2002, Experiment 4), and the outcome of a comparison of the models on these data is shown in Table 2 and Figure 2. In this study, 90 words, varying in frequency, phonological neighborhood size (i.e., the number of words that differ from the target word by a single phoneme), and phonological neighborhood frequency (i.e., the average

Table 1
Parameter Values and Fit Statistics for the Data From Experiment 2, Hulme et al. (1997)

Parameter/Statistic	MPT	CR
i_1	.911	.910
i_2	.718	.725
i_3	.532	.559
i_4	.405	.456
i_5	.165	.264
i_6	.123	.243
i_7	.564	.591
r -high-frequency	.318	.648
r -low-frequency	.100	.500
r^2	.988	.989
G^2	28.39	27.96
BIC	11,927.2	11,926.4
$wBIC$.399	.601

Note—MPT, multinomial processing tree model; CR, constrained Rasch model; i_1 - i_7 , estimates of the i parameter for each serial position; r -high-frequency, estimate of the value of the redintegration parameter for high-frequency words; r -low-frequency, estimate of the value of the redintegration parameter for low-frequency words; G^2 , goodness-of-fit value; BIC, Bayesian information criterion; $wBIC$, Bayesian information criterion model weight.

frequency of the phonological neighbors), were presented to participants in six-item lists. Each participant heard each word once, in each serial position, with the arrangement of the words determined randomly for each participant. The words were all three phonemes in length, and the set of words did not include any that were phonological neighbors. This means that when collapsed across participants, any potential confounds among different words should have been averaged out, and minimal learning of the set of words should have taken place. Also, for each

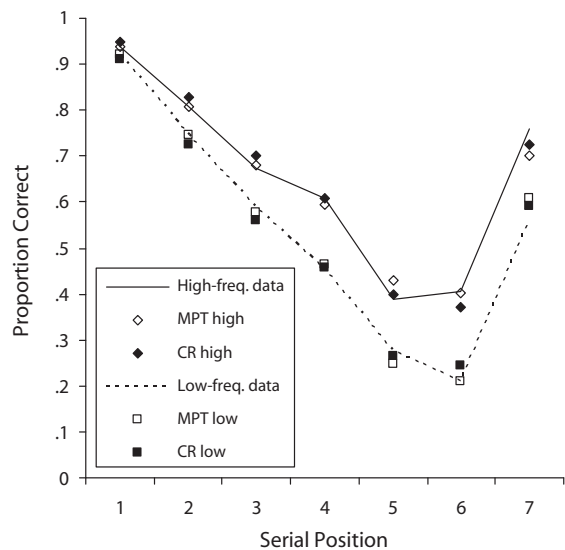


Figure 1. Serial recall data for high- and low-frequency words from Hulme et al. (1997) Experiment 2 (lines), with predicted values from the multinomial processing tree (MPT) model and the constrained Rasch (CR) model.

Table 2
Parameter Values and Fit Statistics for the Data From
Experiment 4, Roodenrys et al. (2002)

Parameter/Statistic	MPT	CR
i_1	.787	.853
i_2	.583	.703
i_3	.370	.540
i_4	.185	.394
i_5	.084	.294
i_6	.403	.567
r -high	.350	.573
r -moderate	.258	.500
r -low	.100	.372
r^2	.975	.994
G^2	31.38	7.74
BIC	7,383.5	7,359.5
wBIC	6.14E-06	1.00

Note—MPT, multinomial processing tree model; CR, constrained Rasch model; i_1 – i_6 , estimates of the i parameter for each serial position; r -high, words with high recall levels; r -moderate, words with moderate recall levels; r -low, words with low recall levels; G^2 , goodness-of-fit value; BIC, Bayesian information criterion; wBIC, Bayesian information criterion model weight.

serial position, the degree of degradation of the different words should have been equivalent, regardless of whether the degradation was due to decay or interference.

To test the fit of the models, three sets of six words that differed on the frequency and neighborhood variables, and consequently in recall level, were selected from the data. Six words were high in frequency and tended to have a large number of neighbors, and these words had the highest recall levels. Another six words were low in frequency, tended to have a small number of neighbors, and were poorly recalled. The remaining six words were of moderate frequency, varied in neighborhood size, and fell between the other two sets in terms of recall performance. These sets will be referred to as *high*, *low*, and *moderate*, respectively.

Each model had eight free parameters and 18 data points to be fitted. As can be seen from Figure 2 and Table 2, the CR model provides a better fit to the data than does the MPT model. The CR model accounts for almost all of the variability, with the r^2 value approaching 1, whereas the MPT model leaves 2.5% of the variability unaccounted for. The G^2 for the CR model is considerably lower than that for the MPT model, as is the BIC value. This difference results in a very small BIC weight for the MPT model, and consequently the ratio of weights between the CR and MPT models results in an index that suggests that the CR is many times more likely to be the better-fitting model.

Testing the Parameter Generality of the Models

One of the major phenomena relating to serial recall is the phonological similarity effect, in which performance is poorer for lists of similar-sounding words than for lists of dissimilar-sounding words. In an early attempt to model the phonemic similarity effect on the recall of letter sequences, Sperling and Speelman (1970) assumed that decay of a short-term trace resulted in the independent loss of phonemes from the trace. In the Sperling and Speelman model, the degraded item could be recalled by guessing from the set of stimuli that matched the intact phonemes in the short-term trace. This is a version of the redintegration notion in which phonological similarity affects the likelihood of correctly redintegrating the item. Schweickert (1993) also described phonological similarity as influencing the redintegration process, although he did so in the context of discussing two redintegrative processes, one lexical and one phonological.

In line with Sperling and Speelman (1970) and Schweickert (1993), we suggest that phonological similarity influences only the likelihood of correctly redintegrating the

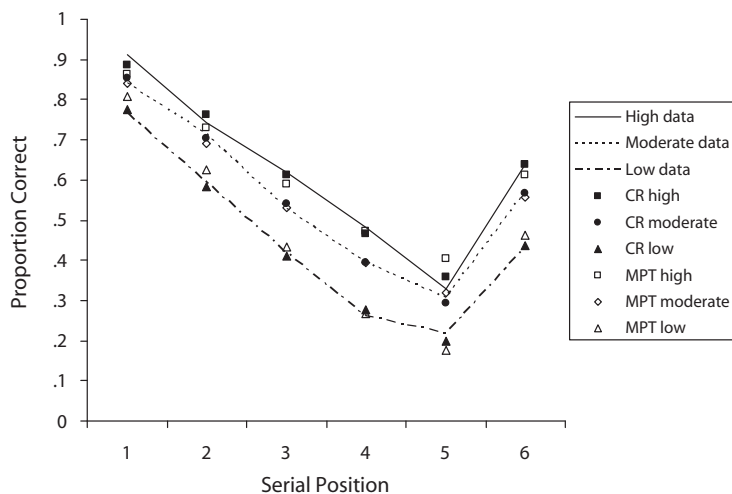


Figure 2. Serial recall data for three sets of words extracted from Roodenrys et al. (2002) Experiment 4 (lines), with predicted values from the MPT model and the CR model.

item. According to the account we outlined above, the outcome of the redintegration process depends on the state of the long-term memory system at the point in time at which the process operates. If a word is presented in a list with words that sound similar to it, then at recall, several words will be active in long-term phonological space close to the target word. If the trace is degraded, this will make it less likely that the correct word will be retrieved. The following experiment was conducted to examine how well the CR model would fit data from phonologically similar and dissimilar word sets by allowing only the redintegration parameter to vary between sets.

There are many phonological similarity experiments in the literature that we might have modeled, but we also wanted to examine another factor from our conceptual account of redintegration. If in the CR model, r relates to how readily retrievable from long-term memory a word is, then r should be unaffected by the length of the list of words to be recalled, assuming that the words are drawn repeatedly from a small set across different lists. That is, r should differ between phonologically similar and dissimilar word sets but not across word lists of different lengths. The following experiment was designed to test how well the CR model could fit a set of data under this constraint in addition to having identical i values across the similarity conditions. The use of different-length lists provides a novel means of testing the generality of the parameters derived from a set of data—by fitting the data from the longest list length and examining how well the parameter values derived from these data fit the data from the other list lengths.

METHOD

Participants

The participants were 47 undergraduate psychology students at the University of Wollongong who completed the experiment as part of a class exercise. Four participants were excluded since they were not native English speakers. The remaining participants ranged in age from 19 to 34, with a mean age of 22 years.

Stimuli

Two sets of eight consonant–vowel–consonant words were selected for use in the experiment. One set contained phonologically similar words (*cake, cane, lake, lane, late, rain, rake, rate*), and the other set contained phonologically dissimilar words (*bomb, deal, park, road, rose, toad, whale, work*). Independent t tests showed that the two sets of stimuli did not differ on word frequency (from the CELEX database; Baayen, Piepenbrock, & van Rijn, 1993), rated concreteness, or imageability (both from the MRC psycholinguistic database; Coltheart, 1981) (all p s > .67).

Procedure

The participants were presented with two blocks of serial recall trials, one for each condition, and the order of testing conditions was assigned randomly. Twenty participants completed the dissimilar condition first. Each block consisted of 24 lists: 6 lists each of three, four, five, or six words drawn at random without replacement from the set of eight words for that condition. List lengths were randomly arranged for each participant. The stimuli were presented on an IBM-compatible PC in 18-point Times New Roman font in black on a white background. Each stimulus was presented for 750 msec, with a 250-msec blank interval between stimuli. Following the final stimulus, a cue (“????”) to recall the list appeared on the screen for

3 sec, followed by the instruction to press the *enter* key to begin the next trial.

The participants were tested in groups of up to 18. Participants wrote their responses on two sheets of 24 lines, with six response spaces indicated on each line. They were instructed to write the words in the order of presentation, beginning with the first word in the list and proceeding in serial order, leaving blanks for any words that they could not recall. They were instructed not to go back and fill in any missing words.

RESULTS

Only responses that correctly matched the position of presentation on the list were scored as correct. The average proportions of correctly recalled words in each serial position in each list length for both the similar and the dissimilar conditions are displayed in Figure 3. The data show the expected pattern: Recall performance decreases overall as list length increases, and performance is better on dissimilar than on similar lists.

The data were subjected to a 2 (condition: similar or dissimilar) \times 4 (list length: three, four, five, or six words) ANOVA, which revealed significant effects of both factors [$F(1,42) = 419.54$, $MS_e = 0.033$, $p < .001$] and [$F(3,126) = 160.81$, $MS_e = 0.007$, $p < .001$]. There was also a significant interaction between the two factors [$F(3,126) = 40.91$, $MS_e = 0.008$, $p < .001$], since the size of the similarity effect decreased as list length increased. The effect of serial position could be evaluated by separate analyses of each length, but it is irrelevant for the present purposes, so this is not presented.

In modeling the data, the r parameter for the similar condition was set to a value that would allow the modeling algorithm to settle on a stable solution relatively easily, and the other values were allowed to vary in the same manner as in the previous simulations. As can be seen from Table 3, the CR model provides a better fit to the data than does the MPT model when all of the list lengths are included (both models using 19 free parameters to fit the 36 data points). The values predicted by the CR model are displayed in Figure 3 as solid shapes. The CR model accounts for 98% of the variability in the data, whereas the MPT model accounts for only 93%. The G^2 for the CR model is considerably lower than for the MPT model, as is the BIC value. This difference results in an extremely small BIC weight for the MPT model, and consequently the ratio of weights between the CR and MPT models results in an index that again suggests that the CR model is many times more likely to be the better fitting model.

The experiment was designed to allow us to examine a different test of how well the models can fit data by examining the generality, or what could be thought of as the reliability, of the parameter values derived from one set of data. The fitting exercise described above shows that the same value of the r parameter in the CR model provides a very good fit to data from lists of different lengths. An alternative way to test this claim is to find the best-fitting values of the parameters for the six-item lists and then use those parameters to fit the data from the other list lengths. If the same r parameter does underlie performance in all

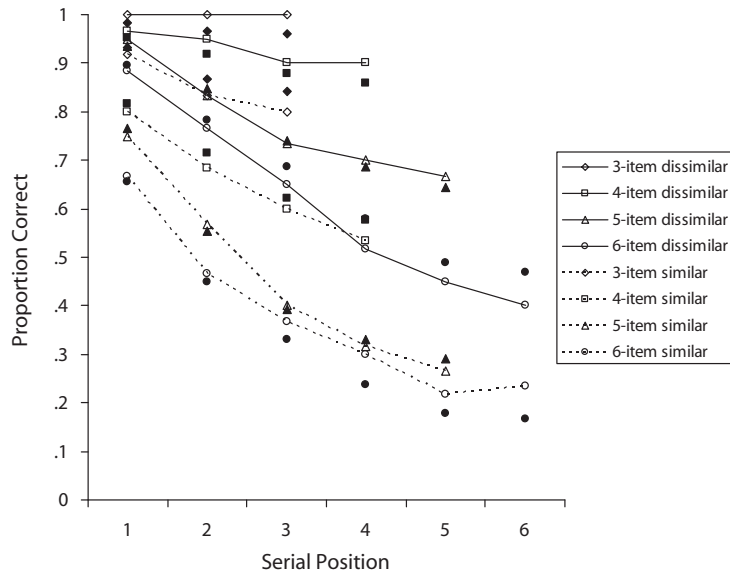


Figure 3. Serial recall data for phonologically similar and dissimilar word sets in lists of varying length (lines and unfilled shapes), with predicted values from the CR model (filled shapes).

list lengths, then the fit obtained under these conditions should not differ greatly from the fit obtained when the data from all the list lengths are fitted at the same time, as shown in Table 3.

The r parameter values for the dissimilar and similar conditions when fitting only the six-item lists are .748 and .500 for the CR model and .412 and .1 for the MPT model. These parameters produce r^2 and G^2 values of .990 and 5.43, respectively, for the CR model, and .943 and 39.30 for the MPT model. When these r parameter values are used to fit the data from all list lengths (allowing the i parameter values to vary), the CR model continues to provide a good fit to the data. The r^2 value for this fit is .970, and the G^2 value is 145.11. In contrast, the fit of the MPT model produces an r^2 value of .914 and a G^2 value of 344.95.

DISCUSSION

The results of the experiment show the typical effects of increasing list length and phonological similarity. The data from this experiment are unusual, perhaps, only in showing a phonological similarity effect for three-item lists. This may be due to the unusual procedure of intermixing list lengths used in this experiment, which meant that participants did not know how long each list would be until the recall cue was presented. Since this is only a minor point and not the focus of the study, it will not be discussed further.

Considering the model fits, it is notable that the CR model provided a much closer fit to the data than did the MPT model in all of the fitting exercises performed. The CR model provided an excellent fit to the total data set and to the six-item lists. It is notable that the goodness of the fit when extrapolating from the six-item lists to the entire

data set is not significantly worse than the best fit for the entire data set shown in Table 3. The same is not true of the MPT model, which performed consistently worse than did the CR model. In addition, it is worth noting that despite the much larger number of data points, the CR model provided nearly as good a fit to this data set as it did to the other two data sets.

It must be noted that the above exercises in fitting the models to the data from this experiment continue a tradition of thinking about redintegration in the context of trace decay models of short-term memory (see, e.g., Hulme et al., 1991; Schweickert, 1993). As such, the effect of phonological similarity has been modeled by varying the r parameter, reflecting the relative state of the item in long-term memory. Items in phonologically similar lists are seen as being less easily redintegrated than are items in dissimilar lists, since very similar items in long-term memory are also highly active, due to their presentation in the experimental lists.

A different account of the effect of similarity on recall and its relation to redintegration has been modeled by Nairne, Neath, and colleagues in their feature model of short-term recall (see, e.g., Nairne, 1990; Neath & Nairne, 1995). The feature model assumes that the encoding of a new item into the short-term store degrades the previous item in the store, with the degree of degradation corresponding to the items' similarity to each other. This means that a list of phonologically similar words would be more degraded at recall than would a list of dissimilar-sounding words, but the level of degradation would not differ across positions in a list. In terms of the CR model, this would translate to each serial position in a condition having the same i value but having different values across different

Table 3
Parameter Values and Fit Statistics for Experiment 1

List Length	Parameter/ Statistic	MPT	CR
3	i_1	.928	.933
	i_2	.855	.867
	i_3	.826	.841
4	i_1	.798	.815
	i_2	.679	.716
	i_3	.562	.621
5	i_4	.498	.576
	i_1	.740	.765
	i_2	.481	.554
	i_3	.253	.392
6	i_4	.149	.331
	i_5	.079	.290
	i_1	.612	.656
	i_2	.340	.450
	i_3	.174	.331
	i_4	.049	.237
	i_5	.000	.178
	i_6	.000	.166
	r -dissimilar	.555	.816
	r -similar	.100	.500
	r^2	.932	.981
	G^2	257.97	96.80
	BIC	8,781.22	8,619.0
	wBIC	6.04E-36	1.00

Note—MPT, multinomial processing tree model; CR, constrained Rasch model; i_1 – i_6 , estimates of the i parameter for each serial position; r -dissimilar, set of dissimilar words; r -similar, set of similar words; G^2 , goodness-of-fit value; BIC, Bayesian information criterion; wBIC, Bayesian information criterion model weight.

conditions. In addition, in the feature model, similarity also affects the likelihood of selecting the correct trace during the redintegration process, and, indeed, this is more important for producing the effect in the feature model.¹ The CR model is not compatible with these assumptions since it produces equivalent recall for each position in a list (since only a single value of the i and r parameters would occur for each condition) and, therefore, fails to produce the most basic characteristic of serial recall performance.

This raises the question of how variables are assigned to particular parameters in the model. This is a particularly important issue for the CR model because it is mathematically symmetrical. When a variable is assigned to one of the parameters, the fit is achieved not only by varying that parameter, but also by allowing the other parameter to vary to produce the best fit. In a model with only two parameters, the same value can be fitted if a change in one parameter is countered by a change in the other parameter. If the model is symmetrical, a variable can be assigned to either parameter as long as the other parameter is free to vary. In the case of this experiment, it is possible to produce exactly the same fit by allowing phonological similarity to influence the i parameter and serial position to influence the r parameter. We will return to this issue later.

GENERAL DISCUSSION

We have demonstrated that, with the same number of free parameters, the CR model provides a better fit to several sets of serial recall data than does the MPT model of

Schweickert (1993) and provides a better measurement model of the redintegration process. Although the CR model is a better mathematical model of this process, the conceptual account on which it is based is similar to that of Schweickert. In the broadest terms, degraded short-term traces are used to retrieve traces from long-term memory. In the CR model, two parameters, i and r , correspond to the state of the traces in those systems, whereas in the MPT model, the two parameters are probabilities of success associated with retrieval from the two stores. What conceptual account of serial recall processes might the CR model be consistent with?

Schweickert originally conceptualized the process of redintegration in terms of speech production. In fact, he suggested that “immediate recall errors are analogous to, if not synonymous with, speech errors” (p. 171) and likened redintegration to monitoring inner speech for errors and correcting them prior to production. Nothing about the CR model contradicts this view, but evidence of phonotactic effects on the recall of nonwords (Thorn & Frankish, 2005) suggests that at least some of the long-term memory contribution to recall is at an unconscious level.

The CR model is equally applicable to theories in which the short-term trace stores a set of phonological representations (see, e.g., Baddeley, 1986; Brown & Hulme, 1995) or an unspecified set of features that may include phonological information (see, e.g., Nairne, 1990). However, Roodenrys et al. (2002) argued that some of these theories cannot account for a facilitative effect of neighborhood size within the redintegration process that they propose. The CR model seems likely to be compatible with the primacy model of Page and Norris (1998b), particularly with a version in which they simulate the phonological output stage of their model using interconnected layers of word and phoneme nodes (Page & Norris, 1998a). Alternatively, Neath (1999) described the primary (short-term) memory store in the feature model as storing cues for the retrieval of information from secondary (long-term) memory. In these terms, the i parameter refers to the integrity of the cues that will be used to access long-term memory, but no characteristic of the model seems to so readily correspond to the r parameter.

Over the years, a number of researchers have argued for a view of short-term memory as being simply that which is currently activated in long-term memory (see, e.g., Cowan, 2001; Norman, 1968). In Cowan’s account, focused attention serves to maintain the activation of items in long-term memory. A slightly different proposal that would also fit with the CR model is that the i parameter relates to the state of a pointer to the item that is maintained by attention and that the r parameter is the state of long-term memory—or, rather, the relative state of the target word in relation to the rest of long-term memory. In the final paragraph of their article outlining their own version of the activated long-term memory view of short-term memory, Ruchkin, Grafman, Cameron, and Berndt (2003) referred to the frontal cortex as providing a pointer system for maintaining the activation of long-term memory structures in the posterior cortex. In these terms, the i parameter corresponds to the integrity of the

pointers held in the short-term/frontal system, and the r parameter corresponds to characteristics of the long-term memory systems. The i parameter might even be regarded as reflecting the activation level of pointers or short-term representations rather than their integrity, as long as activation relates to the likelihood of correct recall.

Lewandowsky and Farrell (2000) modeled the redintegration process using an autoassociative network referred to as a "brain-state-in-a-box" (BSB) network. In their model, they assumed that an associative stage underlies recall prior to redintegration and provides a degraded vector as input to the redintegration process. The state of this vector corresponds to the i parameter in the CR model, and the state of the attractor network in the BSB model corresponds to the r parameter. Lewandowsky and Farrell's model can be seen as a greater specification of the processes that underlie the notion of redintegrability as a characteristic of the long-term trace of the item at the point of recall.

As this discussion illustrates, the CR model is compatible with many theoretical accounts of recall in short-term memory tasks. The minimum conceptual demands of the CR model are a process or state that alters across serial positions, resulting in a decline in performance, and a second process or state that does not vary across serial positions but is influenced by factors such as the preexperimental history of the items. Both Schweickert (1993) and Buchner and Erdfelder (2005) argued that the MPT model could be used to develop theoretical accounts of the processes underlying serial recall. The challenge is to do the same with the CR model, returning us to the issue of how variables are to be assigned to parameters in the model.

It seems to us that the allocation of variables to parameters must be done largely on theoretical grounds; that is, a conceptual model must underlie the application of the model to data. Some manipulations would seem logically applicable to only one parameter, such as the allocation of irrelevant speech to the short-term memory factor, as done by Buchner and Erdfelder (2005) in their simulations using the MPT model. Other manipulations, such as lexicality, logically correspond to the long-term memory factor. However, other manipulations, such as phonological similarity, are more related to conceptual models of serial recall. We have demonstrated that using the assumptions of a trace decay model to determine the allocation of variables to parameters produces an excellent fit to several sets of data. Using a different allocation of variables to parameters corresponds to a different conceptual model; ultimately, which model is favored should not be determined solely by the goodness of fit but also by the plausibility of the conceptual account that it maps onto. For example, the CR model can provide a better fit to the data in Experiment 1 if we assume (1) that a serial position has the same value of i across the different list lengths and conditions and (2) that r varies for each condition and decreases with list length, producing a unique r value for each combination of list length and condition. Such a model uses only 14 parameters (6 serial positions and 8 combinations of list length and phonological similarity) rather than the 19 used in the simulations above, and thus could be consid-

ered more parsimonious. However, this model maps onto a conceptual account in which short-term trace integrity decreases through lists but is unaffected by the overall length of the list, and redintegrability is a function of the length of the list equally for phonologically dissimilar lists as for similar ones. Neither of these aspects of the conceptual account seems particularly plausible.

Before concluding, it is worth returning to the comparison of the CR and MPT models, to consider the differences between them. To some degree, it is to be expected that the CR model will provide a better fit to the data than the MPT model, simply because the CR model simulates a continuous process, whereas the MPT model involves discrete stages. However, these are properties of the conceptual accounts of redintegration behind the two models, not merely of the mathematical instantiation of them. Both mathematical functions model values between 0 and 1, so if the CR model provides a better fit, it lends support to the underlying conceptual account. It is reasonable to see the conceptual account of redintegration described in the introduction as a development of the account underlying the MPT model. Schweickert (1993) suggested that the redintegration process may draw on speech production processes, and we agree with this view. In the MPT model, allowing the probability that the short-term trace is intact to vary with serial position can be seen as a surrogate for a continuous level of degradation, but this is not a strict interpretation of the MPT model. The major difference between our account and that of the MPT model is that we would argue that all items, even if they are intact, undergo redintegration as part of an automatic process inherent in retrieval. The more intact the item is, the less it will be influenced by redintegration and the more likely it is to be recalled correctly.

AUTHOR NOTE

This research was supported by Australian Research Council Grant DP0663642 to the first author. We thank John Dunn and Andrew Heathcote for their help and advice and Gerald Tehan and an anonymous reviewer for helpful comments on an earlier version of this article. Correspondence concerning this article should be addressed to S. Roodenrys, Department of Psychology, University of Wollongong, Northfields Avenue, Wollongong, 2522 NSW, Australia (e-mail: steven_roadenrys@uow.edu.au).

REFERENCES

- BAAYEN, R. H., PIEPENBROCK, R., & VAN RIJN, H. (1993). The CELEX lexical database (Release 1) [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- BADDELEY, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- BROWN, G. D. A., & HULME, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory & Language*, *34*, 594-621.
- BUCHNER, A., & ERDFELDER, E. (2005). Word frequency of irrelevant speech distractors affects serial recall. *Memory & Cognition*, *33*, 86-97.
- COLTHEART, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497-505.
- COWAN, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity (with discussion). *Behavioral & Brain Sciences*, *24*, 87-185.
- CROWTHER, C. S., BATCHELDER, W. H., & HU, X. (1995). A measurement-theoretic analysis of the fuzzy logic model of perception. *Psychological Review*, *102*, 396-408.
- GATHERCOLE, S. E., FRANKISH, C. R., PICKERING, S. J., & PEAKER, S.

- (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 84-95.
- GREGG, V. H., FREEDMAN, C. M., & SMITH, D. K. (1989). Word frequency, articulatory suppression and memory span. *British Journal of Psychology*, **80**, 363-374.
- HULME, C., MAUGHAN, S., & BROWN, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory & Language*, **30**, 685-701.
- HULME, C., ROODENRYS, S., SCHWEICKERT, R., BROWN, G. D. A., MARTIN, S., & STUART, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1217-1232.
- HULME, C., STUART, G., BROWN, G. D. A., & MORIN, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory & Language*, **49**, 500-518.
- LEWANDOWSKY, S., & FARRELL, S. (2000). A redintegration account of the effects of speech rate, lexicality, and word frequency in immediate serial recall. *Psychological Research*, **63**, 163-173.
- MASSARO, D. W. (1987). *Speech perception by ear and eye*. Hillsdale, NJ: Erlbaum.
- MASSARO, D. W., & COHEN, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 753-771.
- MORIN, C., POIRIER, M., FORTIN, C., & HULME, C. (2006). Word frequency and the mixed-list paradox in immediate and delayed serial recall. *Psychonomic Bulletin & Review*, **13**, 724-729.
- MYUNG, I. J., & PITT, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, **4**, 79-95.
- NAIRNE, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, **18**, 251-269.
- NEATH, I. (1999). Modeling the disruptive effects of irrelevant speech on order information. *International Journal of Psychology*, **34**, 410-418.
- NEATH, I., & NAIRNE, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, **2**, 429-441.
- NORMAN, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, **75**, 522-536.
- PAGE, M. P. A., & NORRIS, D. (1998a). Modeling immediate serial recall with a localist implementation of the primacy model. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 227-255). Mahwah, NJ: Erlbaum.
- PAGE, M. P. A., & NORRIS, D. (1998b). The primacy model: A new model of immediate serial recall. *Psychological Review*, **105**, 761-781.
- RASCH, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Paedagogiske Institut.
- ROODENRYS, S. (in press). Explaining phonological neighbourhood effects in short-term memory. In A. S. C. Thorn & M. P. A. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain*. Hove, U.K.: Psychology Press.
- ROODENRYS, S., HULME, C., LETHBRIDGE, A., HINTON, M., & NIMMO, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 1019-1034.
- ROODENRYS, S., & QUINLAN, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory*, **8**, 71-78.
- RUCHKIN, D. S., GRAFMAN, J., CAMERON, K., & BERNDT, R. S. (2003). Working memory retention systems: A state of activated long-term memory (with discussion). *Behavioral & Brain Sciences*, **26**, 709-777.
- SCHWEICKERT, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, **21**, 168-175.
- SPELTING, G., & SPEELMAN, R. G. (1970). Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory* (pp. 152-202). New York: Academic Press.
- STUART, G., & HULME, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 796-802.
- TEHAN, G., & HUMPHREYS, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society*, **26**, 293-296.
- THORN, A. S. C., & FRANKISH, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 729-735.
- THORN, A. S. C., GATHERCOLE, S. E., & FRANKISH, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology*, **50**, 133-158.
- WAGENMAKERS, E.-J., & FARRELL, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, **11**, 192-196.

NOTE

1. That the feature model produces a similarity effect without overwriting of the primary memory features can be demonstrated in an online version of the model (memory.psych.mun.ca/models/feature/) by setting the probability of overwriting to 0 and varying the number of similar features while keeping all other parameters constant.

(Manuscript received February 18, 2007;
revision accepted for publication September 18, 2007.)