Correcting the correction of conditional recency slopes

Simon Farrell

University of Bristol

Author Note

Simon Farrell, School of Experimental Psychology, University of Bristol.

Correspondence should be addressed to Simon Farrell, School of Experimental Psychology, University of Bristol, 12a Priory Road, Clifton, Bristol, BS8 1TU, UK. Email: Simon.Farrell@bristol.ac.uk. Web: http://eis.bris.ac.uk/ pssaf/

Abstract

Farrell (2010) presented some analyses of free recall data that suggest that recency items initially become more accessible as recall progresses, in contrast to the assumptions of temporal drift models. Moran and Goshen-Gottstein (in press) present some challenges to Farrell's (2010) analyses of the change in conditional recency across output position in free recall. Simulations using a very basic free recall model that controls conditional recency across recall show that Farrell's (2010) basic analyses are not substantially biased, while the procedure proposed by Moran and Goshen-Gottstein (in press) introduces a substantial underestimation of the true slopes. The null slopes observed in immediate recall by Moran and Goshen-Gottstein (in press) are not informative of the true slopes characterising the data. Accordingly, Farrell's (2010) results continue to present a challenge to temporal drift models.

Correcting the correction of conditional recency slopes

Farrell (2010) presented a novel analysis of free recall data that posed a problem for unitary models of episodic memory (e.g., Brown, Neath, & Chater, 2007; Howard & Kahana, 2002; Sederberg, Howard, & Kahana, 2008). The recency effect is well established in free recall, and Farrell (2010) was interested in how the tendency to recall the last list item (i.e., the most recently presented information) varies across output position. Unitary models in which recall success is determined by temporal context or temporal distinctiveness—henceforth termed temporal drift models—predict that the last list item will become less recallable over time as it recedes into the past and becomes more confusable with other list items. Directly looking at the probability of recall of the last list item at different output positions is insufficient, as people have a strong tendency to produce only unique items in their recall, implying some response editing process prior to overt recall (Bousfield & Sedgewick, 1944; Davelaar, 2007; Kahana, Dolan, Sauder, & Wingfield, 2005; Murdock & Okada, 1970; Raaijmakers & Shiffrin, 1981). Accordingly, recalling the final list item (henceforth termed the recency item) at one of the first few output positions will make it unavailable for recall at later positions, and this editing process by itself will produce a lower probability of recall of the recency item at later output positions. In order to capture the change in tendency to recall the last item in a manner unaffected by the editing of prior recalls, Farrell calculated the probability of recalling the last list item given that it hadn't yet been recalled. Farrell referred to this measure as *conditional recency*, and obtained a metric for how this measure of recency changes during the act of recall by fitting a linear regression to the relationship between conditional recency and output position. Across a number of free recall experiments conducted by other researchers (Howard & Kahana, 1999; Howard, Venkatadass, Norman, & Kahana, 2007; Murdock, 1962; Murdock & Okada, 1970) Farrell found that this measure increased across initial output positions in immediate recall, but slightly decreased across delayed recall. Temporal drift models models can explain the pattern of conditional recency under delayed recall conditions, but are challenged by the prominent increase in conditional recency observed in immediate recall. By comparing a number of models, Farrell (2010) showed that immediate free recall data—in particular, the increase in conditional recency—were better accounted for by assuming a forward-ordered short-term buffer (see also Farrell, 2012).

Moran & Goshen-Gottstein (2013) have presented a challenge to the analysis of Farrell (2010), arguing that conditional recency is confounded with the overall frequency of recall of the last item (they term the overall probability of recall of the final list item *nominal recency*). Essentially, Moran and Goshen-Gottstein's concern is that the higher levels of recall under immediate recall give a larger change in conditional recency—compared to delayed recall—even when the change in recency across output positions is apparently controlled. Moran and Goshen-Gottstein (2013, p. 11–12) give the example where only two items are recalled per trial. If the overall recall probability of the recency item is 30%, and assuming that the recency item is equally likely to be recalled at either output position, the slope relating conditional recency to output position is 0.027. In contrast, if the overall probability of recall of the last item is 80% (such that the recency item has a 40% chance of occurring at each of the two output positions), the conditional recency slope becomes 0.2667.

Motivated by these concerns, Moran & Goshen-Gottstein (2013) applied a correction factor in which the obtained conditional recency slope (that is, the slope relating conditional recency to output position) is considered against a baseline distribution of slopes obtained by Monte Carlo simulation. The slopes in the distribution are generated by repeatedly shuffling the order of recalled items in output positions 1–4 (the range of output positions analysed in the relevant section of Farrell, 2010), and calculating a slope for each shuffled version of the data (see Moran & Goshen-Gottstein, 2013, for further details). The justification for this procedure is to control nominal recency, whilst disrupting any output ordering evident in the original unshuffled data. Although their results for delayed recall agree with those of Farrell (2010) in showing a null or negative slope, Moran & Goshen-Gottstein (2013) failed to find any overall trend to a positive slope in immediate recall; indeed, of the 9 immediate recall conditions examined, five gave numerically negative slopes when compared to the Monte Carlo samples.

Below, I address the concerns raised by Moran & Goshen-Gottstein (2013) and the correction those authors proposed. Briefly, Moran and Goshen-Gottstein's (2013) concerns about a confound between conditional and nominal recency are look compelling, but actually have no effect on the slopes measured using Farrell's (2010) method. In contrast, the correction factor introduced by Moran & Goshen-Gottstein (2013) is inappropriate, as it heavily biases the conditional recency measure downwards. As discussed below, the major issue with Moran and Goshen-Gottstein's (2013) conceptual and statistical analysis is that it does not take into account the response editing that originally motivated Farrell to analyse conditional recency, and ignoring this assumption leads Moran and Goshen-Gottstein (2013) to overestimate the conditional recency generated under random output ordering (the assumed null model).

Are Conditional Recency Slopes In Immediate Recall Upwardly Biased?

The core focus of Moran and Goshen-Gottstein's (2013) critique is on the slope relating output position to conditional recency; as noted by Farrell (2010), the function is not necessarily linear, but fitting a linear function at least gives us some idea of the general trend in recall of the terminal list item across output positions. In Farrell (2010), these were only examined across the first 4 output positions, because in many data sets conditional recency fell at further output positions, consistent with a model assuming a short-term buffer (Farrell, 2010) or something similar (Farrell, 2012).

The major discrepancy between the results of Farrell (2010) and Moran & Goshen-Gottstein (2013) is in the results for data sets where immediate free recall was required. Here, Farrell found substantially positive slopes, while Moran &

Goshen-Gottstein (2013) found no systematic upwards trend. Moran and Goshen-Gottstein's (2013) conclusion is there is a positive bias in the slopes calculated by Farrell (2010). However, in order to make any strong claims about bias in either method, what is needed is some benchmarking of the methods against a known reality. We can determine whether the conditional recency slopes reported by Farrell (2010) truly are biased by simulating data from a very basic model in which the tendency to recall the last item across recall (independent of response editing) is controlled. Accordingly, we have perfect knowledge of the process generating the data, and compare "truth" to the results given under the two types of analysis.

Data were simulated here for a trial by stepping through successive output positions j, and at each output position sampling a list item. For each j, the probability of recall of the recency item $p_{end}(j)$ was calculated as

$$\begin{cases} c + \beta(j-1) & \text{if recency item available,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

where parameters c and β are a constant and slope of the simulated conditional recency function. The probability of recall of non-terminal items was obtained by evenly dividing the probability mass $1 - p_{end}$ amongst non-terminal items that hadn't yet been recalled. Data were only simulated for $j = 1 - 4$, as the conditional recency slope analyses focussed on only these positions (Farrell, 2010; Moran & Goshen-Gottstein, 2013).

The output of the simulation is a set of simulated data coded in the same format as the analysed data (Howard & Kahana, 1999; Howard et al., 2007; Murdock, 1962; Murdock & Okada, 1970). Specifically, the simulation produces a $T \times S$ matrix, where $T = 20$ is the number of simulated trials, and $S = 10$ is the list length being simulated, and each element of the matrix codes the item that was recalled at that position on that trial. For a particular trial, this simple model first determines whether the last item was recalled at the first output position (with probability $p_{end}(1)$) or not. If it was, the code for the last list item is entered (so for list length 10, this would be "10"), and if not, the code for the item that was recalled is entered (e.g., "3" for the 3rd list item). The model then proceeds to the second position. If the last item wasn't recalled at the first output position, it is recalled at the second output position with probability $p_{end}(2)$; in all other cases, another item was randomly selected for recall (again, only those items that hadn't yet been recalled). This process was repeated for output positions 3 and 4, and the remainder of the matrix was filled with a dummy value of 0; the analyses did not include data beyond output position 4. The simulation results are reported from 1000 simulation runs, and for each simulation run Moran and Goshen-Gottstein's (2013) Monte Carlo method was based on 1000 reshuffles of the simulated data.

The β parameter in Equation 1 represents the simulated relationship between output position and the tendency to recall the last list item (i.e., the change in conditional recency). On top of that, both parameters β and c will contribute to nominal recency, as they are both positively related to the probability of recalling the last list item. In the simulations, β was varied across the values {0, 0.05, 0.1, 0.15, 0.2}, and the constant c took on the value 0.0 or 0.3 to reflect a reasonable range of conditional recency for the first output position seen in Farrell (2010). Of issue is whether conditional recency slopes calculated from the simulated data accurately recover the β used to generate the data, even in light of variations in nominal recency.

Figure 1 shows the results of the simulation. For reference, the top-left of the figure shows the true value of β fed in to the model. The top-right panel shows the slopes estimated using Farrell's (2010) basic slope analysis. The values obtained from Farrell's method very closely track the true slopes, with no suggestion that the slopes have been overestimated. For comparison, the bottom-left panel shows the the resulting slopes when these slopes are modified by subtracting away the mean slope across samples from Moran and Goshen-Gottstein's (2013) Monte Carlo method. The slopes no longer track the true values; in particular, the slopes are substantially underestimated for higher levels of recall, especially when recall is bolstered by a high constant in the conditional recency function. For comparison to Moran & Goshen-Gottstein (2013), the bottom-right panel plots the same results as z-scores (by dividing the differences plotted in the bottom-left panel by the standard deviation of the Monte Carlo generated slopes; Moran & Goshen-Gottstein, 2013), and again fails to track the true slopes.

This analysis indicates that there is no cause for concern about bias in the conditional recency slopes for immediate recall estimated in Farrell (2010), as that method of analyses accurately recovers the true β in the data. Critically, the results also show no evidence that nominal recency confounds attempts to estimate conditional recency. As noted above, variations in c and β produce variations in nominal recency across the points in Figure 1. Specifically, for $c = 0$, the expected nominal recency for the points (from left to right) is 0, 0.27, 0.50, 0.67, and 0.81; and for $c = 0.3$, the values are 0.76, 0.85, 0.92, 0.96, and 0.99. If Moran and Goshen-Gottstein claims of a confound are correct, the difference in nominal recency for different values of c should have produced a difference in conditional recency (i.e., a separation of the two lines in Figure 1); this was not observed.

The implication is that the null slopes seen under Moran and Goshen-Gottstein's (2013) correction in their paper are due to a strong tendency, seen here in the bottom row of Figure 1, for that correction to return a diminished slope, even in (indeed, especially in) cases where slopes are substantially positive. Accordingly, we are limited in the inferences we can make when a null or negative slope is observed in empirical data using Moran and Goshen-Gottstein's (2013) suggested method.

This may seem rather puzzling, as Moran and Goshen-Gottstein's (2013) Figure 4 seems to show that Farrell's (2010) "uncontrolled" slopes will produce a positive slope for larger values of nominal recency even when the assumed slope is 0. However, the difference between the analysis here and that presented in Moran & Goshen-Gottstein (2013) is in *what* is being controlled. In the simulation here, we controlled the preference for recalling the last item across output positions, taking into account response editing, by specifying different known values of β. In contrast, Moran and Goshen-Gottstein's (2013) theoretical analysis and Monte Carlo procedure control the *unconditional* probability of recall of the last item. In other words, by shuffling the order of items, it is assumed that the last list item is equally likely to occur at any of the first four output positions examined. Although this sounds reasonable, it neglects to consider the actions of response editing that originally motivated Farrell to examine conditional recency.

Consider a simple chance model where the conditional probability of recalling the last item is a constant 0.5. The probability of recalling the last list item at the first output position will be 0.5. The probability of recalling the terminal item at the second output position will be the conditional probability of recall that item (0.5) multiplied by the probability that the item hasn't yet been recalled (also 0.5), to give a recall probability for 0.25. For output positions 3 and 4, the resulting raw recall probabilities are 0.125 and .0625. Due to the actions of the assumed response editing, the unconditional probability of recalling the last item decreases, although that item continues to be uniformly competitive for recall if it in remains unrecalled.

In contrast, by assuming a constant unconditional probability of recalling the terminal item, Moran and Goshen-Gottstein (2013) assume an increasing tendency to recall the terminal item that exactly counteracts the effects of response editing. For example, for a nominal recency of 0.8, the unconditional probability of recall of the terminal item under their shuffling procedure is 0.2. At the first output position, this equates to a conditional recall probability of 0.2 across output positions. At the second output position, the conditional recall probability must be $0.2/0.8 = 0.25$ in order to produce an unconditional probability of recall of 0.2. At positions three and four, the conditional recall probabilities must be equal to $0.2/0.6 = 0.33$ and $0.2/0.4 = 0.5$ in order to fix the unconditional probabilities at 0.2. Accordingly, the reshuffling procedure introduces a strong positive trend into the conditional recall probabilities, and when subtracted from any trend in the data often produces a null or negative difference in slopes (bottom right of Figure 1).

The difference between the two approaches can be compared to an example more familiar to experimental psychologists. Imagine we wish to randomize the time between the onset of two events in an experiment (e.g., presentation of a stimulus and presentation of a signal to respond). Although we might sample the times from a uniform distribution, a more appropriate distribution is an exponential distribution (Luce, 1986). An exponential distribution has a constant hazard function, meaning that the monetary expectation that the stimulus will appear (i.e., in the next arbitrarily small slice of time) is independent of how much time has passed in the trial. This is exactly the situation assumed in Farrell's conditional recency analysis, where the passage of time is replaced by recall events, and the tendency for the signal to appear is replaced by the tendency for the last list item to be recalled. In contrast, a uniform distribution gives a linearly increasing probability of observing the signal as time progresses. In a similar vein, Moran and Goshen-Gottstein's analysis random ordering of outputs controls overall expectancy of the final list item, but does not control its momentary tendency to be recalled across output positions.[1]

Implications for delayed recall

The preceding simulations showed that—Moran and Goshen-Gottstein's (2013) conceptual analysis notwithstanding—the slope analyses reported by Farrell (2010) are not substantially biased by nominal recency. This also has implications for considering the dissociation between immediate and delayed recall. Farrell (2010) showed that delayed recall slopes tended to hover around 0, and were even slightly negative for continuous

[1]One caveat here is that recall events are heavily driven by sampling from the presented list items. Accordingly, if the final list item hasn't been recalled yet (i.e., at a particular output position), it is more likely to be recalled in future given that another list item will have been sampled in its place, and taken out of the pool of items yet to be overtly recalled. Farrell (2010) accounted for this artefact in two ways: along with empirical slopes he also plotted slopes from a chance model assuming that the last list item was as likely to be recalled as any other list item, but assuming response editing; and his modelling assumed a diminishing pool of items were sampled for recall as successively recalled items were removed from that pool.

distractor recall. In some unpublished simulations mirroring those reported in Figure 1, it was found that varying nominal recency did not produce an artefactual difference between immediate and delayed recall. Indeed, this can trivially be shown analytically. Assume probabilities $p_1 + c$ and $p_2 + c$ of recalling the recency item at the first and second output positions (given that it is available), with the constant c allowing us to vary nominal recency without affecting the slope. Assuming no repetitions, the observed unconditional probability of recall will be $p_1 + c$ at the first output position, and $(1 - (p_1 + c))(p_2 + c)$ at the second. The observed conditional probabilities will be $p_1 + c$ and $p_2 + c$ (i.e., the true underlying probabilities will be recovered), and the observed conditional recency slope will be $(p_2 + c) - (p_1 + c) = p_2 - p_1$, so that the slope is not affected by additive changes to p_1 and p_2. Accordingly, the null or negative values observed in delayed recall by Farrell (2010) appear to be unbiased. Of course, varying only the slope will vary nominal recency, but thinking of overall recency as a confounding limiting factor is misleading; rather, the total number of items recalled is instead an aggregate measure of recall, and will be determined by the processes occurring at each recall step, including those related to any change in the tendency to recall the recency item across output positions.

Before moving on, note that even we did suspect that Farrell's (2010) analysis produced an artificial difference between immediate and delayed recall (based, for example, on the difficulty of interpreting single dissociations; Teuber, 1955), the only other possible interpretation of the results is that a positive conditional recency slope is seen in both immediate *and* delayed recall. This would pose as much of a challenge to unitary models—if not more—than the dissociation suggested by Farrell (2010), as those temporal drift models are strongly inclined to produce negative slopes irrespective of delay.

Closing comments

It should be emphasised that the slopes forming the core focus of Moran and Goshen-Gottstein's (2013) critique represent only one part of the analysis reported by

Farrell (2010). More compelling evidence comes from his explorations of the entire conditional recency function. As noted by Farrell (p. 326), fitting a linear regression to the conditional recency function is only useful for obtaining a general idea of the trend in the data. The heuristic nature of this device is particularly obvious in comparisons Farrell (2010) made between different immediate recall data sets, where some show an inverse U-shaped function, while others show a shallow drop before "kicking up" at the last legitimate output position.

In addition, Farrell (2010) stressed the importance of model selection for addressing the role of different mechanisms in free recall performance. Moran and Goshen-Gottstein (2013) spend some time (pp. 4–5) arguing that any demonstration of a weakness in Farrell's (2010) analyses *a priori* questions his modelling results: 'it is reasonable to assume that had the confound been uncovered "in time", then data fitting of the null effect would not have been undertaken in the first place, let alone survived the peer-review process' (p. 5). This misses the critical point that in Farrell's (2010) modelling, the data and the model's output were both subjected to exactly the same analysis in order to extract conditional recency slopes. Accordingly, any bias in the calculation of the slopes should have affected both the empirical and theoretical estimates. In addition, much of the weight in Farrell's (2010) conclusions was carried by comparison of the (maximum likelihood) fit of the models to individual responses; Farrell (2010) showed that that fit of the temporal drift models was substantially improved by the addition of an ordered buffer that drove recall over the initial few output positions.

Moran and Goshen-Gottstein (2013) rightly note that some more complicated variants of temporal drift models may be able to account for the positive conditional recency slopes seen in Farrell (2010). Moran and Goshen-Gottstein (2013) did not offer any specific suggestions for a mechanism that could overcome the tendency of these models to produce a negative slope, and the onus is on those who argue for temporal drift models to demonstrate that their models can account for the data (and that any such ability is not simply due to the added flexibility in more complicated models). Farrell's (2010) data were simply offered up as a compelling factor that potentially discriminates between various theoretical viewpoints, and are theoretically diagnostic because the predictions of temporal drift models follow from core properties of those models. In addition, Farrell's (2010) results need not be taken as evidence against all unitary models. Farrell (2012) accounted for patterns of conditional recency in a unitary model by assuming that all recall is driven by recall of groups of information, and that information within a group is recalled in an ordered fashion. Farrell's (2012) model produces substantial conditional recency in immediate recall because the context used to access the last group is carried over into recall rather than needing to be (noisily) recalled; accordingly, if the last group is recalled first, the superior quality of the group context that is used to cue the items in that group makes the last group behave very much like a short-term ordered buffer, without assuming a separate short-term store.

Finally, given that Moran and Goshen-Gottstein's (2013) reshuffling procedure seems reasonable, are there any situations where it would be appropriate? One obvious case is where recall follows the random sampling model described in Equation 1 (which in turn mimics standard random search models and contemporary free recall models), but where participants do not edit repetitions out of their responses (Bousfield & Rosner, 1970; Kahana et al., 2005). However, in this case an examination of conditional probabilities as in Farrell (2010) does not seem necessary, and unconditional probabilities could be directly regressed on output position. Moran and Goshen-Gottstein's (2013) procedure would also be appropriate in situations where recall was not generated by a random search model, but where the entire recall sequence was assembled before being output, rather than being emitted piecemeal. For example, Dennis (2009) assumed that serial recall occurs by calculating the posterior probability of entire recall sequences; in this framework, repetitions can be minimized or prevented by down-weighting the prior probability of sequences containing repetitions. Such a model would certainly be feasible, but would be incompatible with the models targeted by Farrell's (2010) critique, which explicitly assume changes in the accessibility of items during the act of recall (Brown et al., 2007; Howard & Kahana, 2002; Sederberg et al., 2008).

References

Bousfield, W. A., & Rosner, S. R. (1970). Free vs. uninhibited recall. *Psychonomic Science*, *20*, 75–76.

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology*, *60*, 225–233.

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

Davelaar, E. J. (2007). Sequential retrieval and inhibition of parallel (re)activated representations: a neurocomputational comparison of competitive queuing and resampling models. *Adaptive Behavior*, *15*, 51–71.

Dennis, S. (2009). Can a chaining model account for serial recall? In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the thirty-first annual meeting of the cognitive science society* (pp. 2813–2818). Austin, TX: Cognitive Science Society.

Farrell, S. (2010). Dissociating conditional recency in immediate and delayed free recall: A challenge for unitary models of recency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 324–347.

Farrell, S. (2012). Temporal clustering and sequencing in working memory and episodic memory. *Psychological Review*, *119*, 223–271.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 923–941.

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

Howard, M. W., Venkatadass, V., Norman, K. A., & Kahana, M. J. (2007). Associative processes in immediate recency. *Memory & Cognition*, *35*, 1700–1711.

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: age differences in editing of overt responses. *Journal of Gerontology: Psychological Sciences*, *60B*, P92–P97.

Luce, R. D. (1986). *Response times*. Oxford: Oxford University Press.

Moran, R., & Goshen-Gottstein, Y. (2013). The conditional-recency dissociation is confounded with nominal recency: should unitary models of memory still be devaluated? *Psychonomic Bulletin & Review*.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488.

Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Experimental Psychology*, *86*, 263–267.

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

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

Teuber, H.-L. (1955). Physiological psychology. *Annual Review of Psychology*, *6*, 267–296.

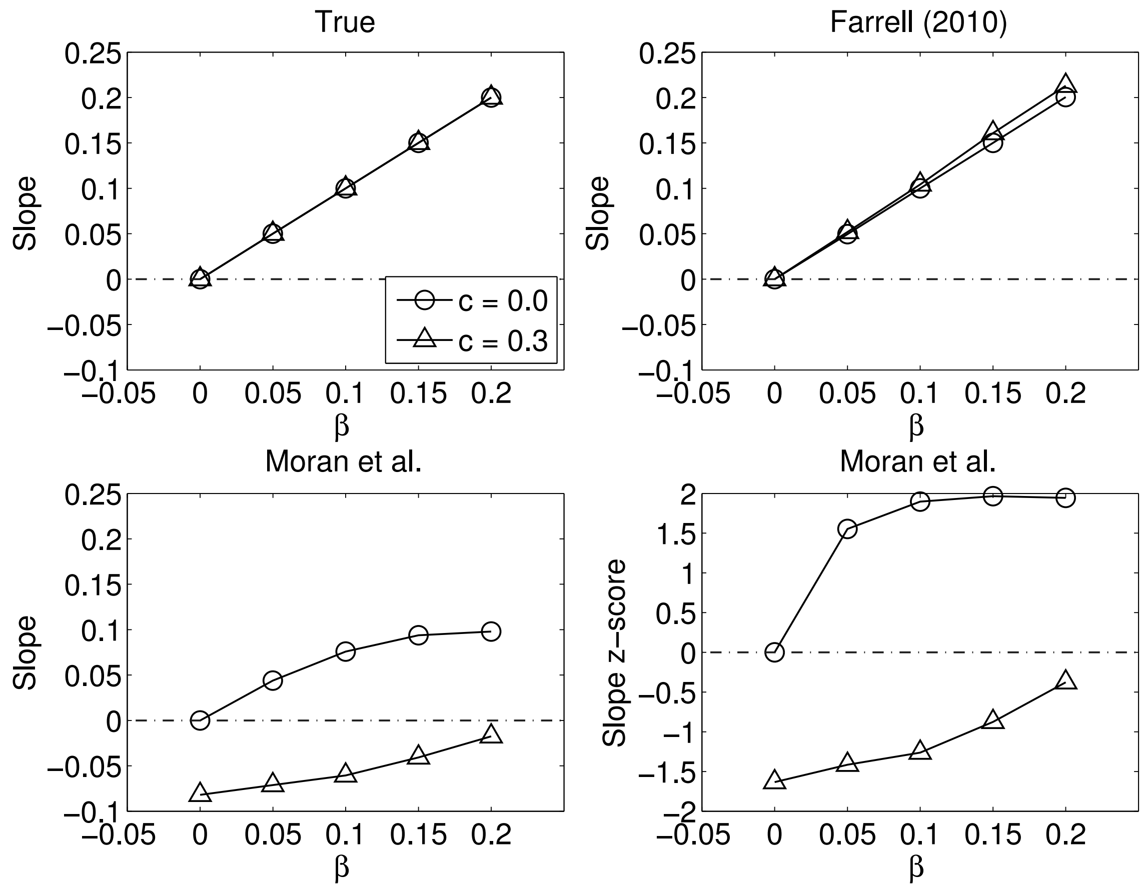

Figure 1. Simulated conditional recency slopes in immediate free recall. Top left: "true" slopes being estimated. Top right: Slopes estimated using Farrell's analysis. Bottom left: Slopes from top-left panel after subtracting the mean of the slopes from Moran and Goshen-Gottstein's Monte Carlo procedure. Bottom-right: Slope estimates from Moran and Goshen-Gottstein's full correction.