Under review (2^{nd} round) at the Journal of Experimental Psychology: General

Individual Differences in Memory Search and Their Relation to Intelligence

M. Karl Healey, Patrick Crutchley, and Michael J. Kahana
University of Pennsylvania

Attempts to understand why memory predicts intelligence have not fully leveraged state-of-the-art measures of recall dynamics. Using data from a multi–session free recall study we examine individual differences in measures of recall initiation and post–initiation transitions. We identify four sources of variation: one corresponding to primacy, another to recency, a temporal factor corresponding to transitions mediated by temporal associations, and a semantic factor corresponding to semantically–mediated transitions. Together these four factors account for 83% of the variability in overall recall accuracy, suggesting they provide a nearly complete picture of recall dynamics. We also show that these sources of variability account for over 80% of the variance shared between memory and intelligence. The temporal association factor was the most influential in predicting both recall accuracy and intelligence. We outline a theory of how controlled drift of temporal context may be critical across a range of cognitive activities.

Complex behavior such as having a conversation, reading a paper, or making a decision relies on the coordinated operation of many cognitive processes. For over 100 years psychologists have attempted to understand how such coordination is achieved. One of the earliest findings was that performance on simple memory span tasks predicts success on more complex tasks (Jacobs, 1887). Dozens of studies have since confirmed that span performance correlates with a wide range of cognitive abilities (for meta-analyses see Ackerman, Beier, & Boyle, 2005; Daneman & Merikle, 1996). In trying to understand the connection between memory and complex cognition, the literature has come to focus on general intelligence, as it constitutes a theory-neutral statistical factor that contributes to almost all cognitive tasks (i.e., the "positive manifold"; Carroll, 1993). The question of which memory processes are critical in predicting intelligence has animated the individual differences literature for over 30 years (Daneman & Carpenter, 1980; Mogle, Lovett, Stawski, & Sliwinski, 2008; Turner & Engle, 1989; Unsworth, Brewer, & Spillers, 2009).

Most of this work has focused on limitations in working memory and attention (e.g., Hasher, Lustig, & Zacks, 2007; Kane, Conway, Hambrick, & Engle, 2007; Oberauer, 2002; Towse, Hitch, & Hutton, 1998). The focus on working memory likely has roots in the fact that span tasks were developed to measure the ability to hold information active in primary memory rather than to measure the ability to retrieve information from secondary memory (Jacobs, 1887). As the idea of a passive primary memory matured into the notion of a *working* memory system that both stores and manipulates information (Baddeley, 2003; Miyake & Shah, 1999), new "complex" span tasks were designed that required simultaneously storing and processing of information (Daneman & Carpenter, 1980; Turner & Engle, 1989). These complex span tasks have proven to be even better predictors of intelligence than simple span tasks, further solidifying the central role of working memory in the search for the link between intelligence and memory. Despite extensive investigations, a consensus on which processes are critical has failed to emerge (for a variety of competing perspectives see Conway, Jarrold, Kane, Miyake, & Towse, 2007).

Recent evidence suggests that part of the difficulty is that in addition to working memory, episodic memory also contributes to the correlation between span and intelligence. Healey and Miyake (2009) found that span tasks require considerable attentional resources during retrieval, which is inconsistent with the view that items are held in working memory and easily accessible. Mogle et al. (2008) and Unsworth et al. (2009) have shown that after accounting for variation in episodic tasks such as free recall, paired associate learn-

This research was funded by National Institutes of Health Grant MH55687 and National Science Foundation Grant NSF1058886. We thank Jonathan Miller for assistance with designing and programming the experiment, and Kylie Hower, Joel Kuhn, and Elizabeth Crutchley for help with data collection.
Correspondence concerning this article should be addressed to M. Karl Healey (healeym@sas.upenn.edu) or Michael J. Kahana (kahana@psych.upenn.edu) at University of Pennsylvania, Department of Psychology, 3401 Walnut St., Room 303, Philadelphia.

phone: (215) 746–3500
Version of September 26, 2013
8827 Words (Approximate due to use of LaTeX)

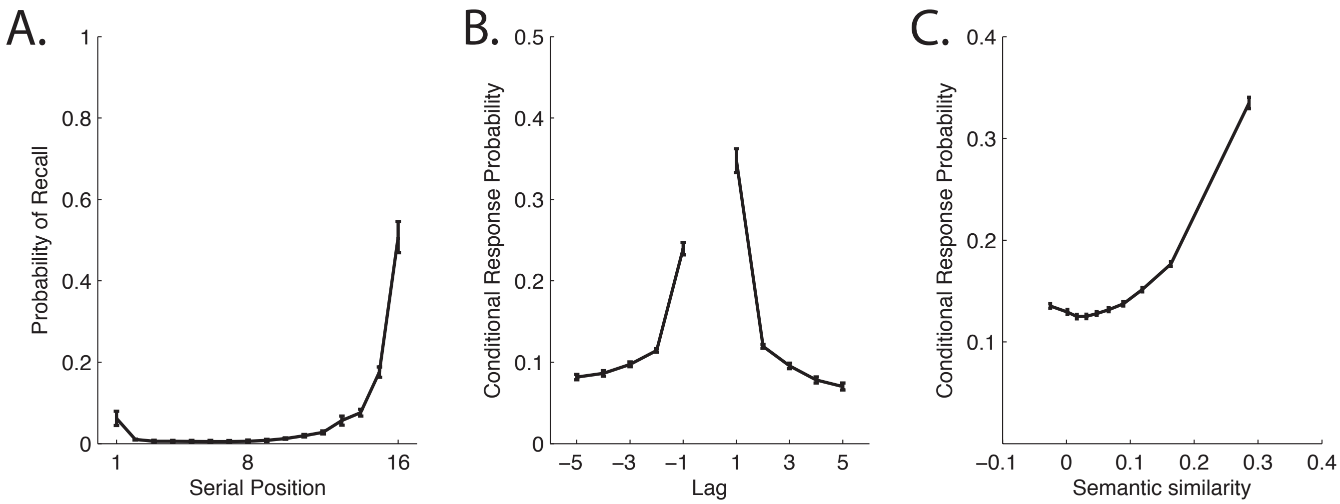

Figure 1. The Recall Dynamics Functions. Data are from the 141 participants who completed Experiment 1 of the Penn Electrophysiology of Encoding and Retrieval Study. Participants studied multiple 16–item lists for immediate free recall (see the Methods section for details). **A**. Probability First Recall curve. **B**. Lag-Conditional Response Probability curve. **C**. Semantic-Conditional Response Probability curve. See the text for details on how these curves are computed. Error bars are 95% within-subject confidence intervals (Loftus & Masson, 1994).

ing, and prose recall, the correlation between span and intelligence is either eliminated or considerably reduced. Discovering which episodic memory processes are related to intelligence is now a priority for individual differences research (Ratcliff, Thapar, & McKoon, 2011; Unsworth, Brewer, & Spillers, 2013).

Prior to the current focus on working memory, some early individual difference work examined episodic memory tasks (e.g., Carlson, 1937; Christal, 1958; Games, 1962; Kelley, 1964; Underwood, Boruch, & Malmi, 1978). This work attempted to understand the relationships among both span and classic episodic memory tasks (e.g., free recall, paired associates, recognition) by deriving an overall summary measure for each task, such as overall recall accuracy, and examining the correlations among the summary measures. The main conclusion was that although there were identifiable sub-groups of memory tasks (e.g., span tasks versus associative tasks), the tasks also loaded onto a common factor (for reviews see Beier & Ackerman, 2004; Kane & Miyake, 2008). Although this older literature clearly established that episodic memory tasks share common variance, perhaps due to the focus on summary measures, it has not shed much light on which memory processes underlie the correlation with intelligence.

Proceeding largely in parallel to the individual difference literature (Carroll, 1993; Cronbach, 1957; Kane & Miyake, 2008; Underwood, 1975), the experimental study of episodic memory has focused not on the correlations of summary measures across tasks but on developing a detailed understanding of the cognitive processes at work within particular tasks. This work has provided a set of sophisticated measures of recall dynamics, which have only recently begun to inform the individual difference literature (e.g., Healey & Kahana, in press; Sederberg, Miller, Howard, & Kahana, 2010; Unsworth, 2009). Here we examine individual differences in recall dynamics in an effort to illuminate the correlation between memory and intelligence. We begin by reviewing the dynamics of memory search.

The Dynamics of Memory Search

The dynamics of memory search can be decomposed into recall initiation and post–initiation transitions. Probability of first recall (PFR) curves (Figure 1A) measure initiation by showing the probability of initiating from each serial position (Hogan, 1975; Howard & Kahana, 1999; Laming, 1999). In immediate free recall of supra-span lists (Grenfell-Essam & Ward, 2012), participants tend to initiate from the last serial position (Deese & Kaufman, 1957).

Post-initiation dynamics are revealed by the order in which items are recalled. Both long–standing semantic associations and newly formed episodic (temporal) associations exert a powerful influence on recall order. The influence of temporal associations can be described by how the probability that recall of item i is followed by recall of item j changes as a function of the distance, or lag, between i and j in the original list. For example, if $i = 5$ and $j = 6$ we would have a lag, $j - i$, of $+1$. Plotting these probabilities for a range of lags gives a lag-CRP (Conditional Response Probability) function. Lag-CRPs are computed by dividing the number of times a transition of a given lag was *actually* made by the number of times it *could* have been made (Kahana, 1996). Lag-CRPs (Figure 1B) show a strong temporal contiguity effect.

We can examine the influence of long–standing semantic associations on transition probabilities (Bousfield, 1953; Romney, Brewer, & Batchelder, 1993) using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), which measures the proximity of words in a multidimensional model of semantic space. Using LSA values to create a semantic-CRP curve (Figure 1C) reveals a strong semantic contiguity effect (Howard & Kahana, 2002). Together the three curves (Figure 1), which we call the Recall Dynamics Functions, provide a summary of the dynamics of memory search.

We have recently shown that these three Recall Dynamics Functions exhibit a remarkable level of qualitative consistency across individuals (Healey & Kahana, in press). The functions do, however, show individual differences. Whereas most participants tend to initiate recall with the very last item of the list, a sub–group of participants tends to initiate from a few items back from the end, and another tends to initiate from the beginning of the list. Post–initiation dynamics also show variation, but it is quantitative variation in level of the functions, not their shape. Here we provide a more detailed examination of these individual differences and their relation to intelligence.

Individual Differences in Memory Search

The fundamental goals of studying individual differences in memory and intelligence are to identify which processes show variation across individuals, and which of those processes are correlated with IQ. A common approach in the individual differences literature has been to identify a cognitive process that may be critical in both memory and intelligence tasks. Researchers then administer several tasks designed to measure memory ability, intelligence, and the putative mediating process and derive a summary measure from each task. These summary measures are then used to extract a latent variable, which is taken as a index of the underlying construct. An assumption here is that although the putative underlying construct contributes to all three types of tasks, the latent variable extracted from the third set of tasks provides a purer measure of that process. Researchers can then test if statistically controlling for variation on that new, purer, measure eliminates (i.e., mediates) the correlation between memory performance and IQ. This logic has been applied to testing the idea that episodic memory processes explain the correlation between memory tasks and IQ. For example, Mogle et al. (2008) showed that the correlation between span tasks and fluid intelligence was no longer significant when variation in episodic memory tasks, including free recall, was controlled for. Unfortunately, this approach has failed to provide consistent results. For example, Unsworth et al. (2009) have shown evidence for partial but not complete mediation.

One reason for this difficulty may be that the mediation approach attempts to infer the internal factor structure of memory from the pattern of correlations among summary measures derived from a variety of different memory and non–memory tasks. As such, the approach is highly theory driven in that the researcher must have some prior hypothesis about which processes are critical. The need for prior theories is, in part, due to the reliance on summary measures. Summary measures reflect the final outcome of all the contributing cognitive processes. In other words, they take the output of multiple processes and compress them into a single number. This compression makes it difficult to directly determine how many processes contributed to the summary measure, forcing the researcher deduce the underlying processes from correlations among different tasks.

The Recall Dynamics Functions, however, allow us to take a more data–driven approach to studying individual differences. Instead of inferring the internal factor structure of memory from correlations across different tasks, we can examine detailed measures of performance derived from a single task. Essentially, the recall dynamics functions uncompress the summary measure of overall recall, providing a window on the cognitive processes that produce overall performance.

Some important facts about individual differences in recall dynamics have already been discovered: temporal and semantic contiguity seem to be universal principles across individuals (Healey & Kahana, in press), temporal contiguity is positively correlated with overall recall (Sederberg et al., 2010; Spillers & Unsworth, 2011), and individuals who exhibit both strong primacy and strong recency effects tend to have higher fluid intelligence scores than individuals who exhibit either weak primacy or recency (Unsworth, Brewer, & Spillers, 2011). Although important first steps, these studies have not fully embraced the idea of moving beyond summary measures to examine the full richness of the Recall Dynamics Functions and have instead used summary measures such as temporal contiguity scores and latency to first recall (e.g. Sederberg et al., 2010; Unsworth, 2009).

The utility of summary measures is that they provide a single variable on which to compare individuals. By contrast, the Recall Dynamics Functions of Figure 1 include 36 separate variables (16 points in the PFR, and 10 points each in the lag-CRP and semantic-CRP). Directly examining individual differences on these 36 dimensions would be intractable. Instead we use factor analysis as a tool to reduce the Recall Dynamics Functions to a manageable number of dimensions while retaining the richness of the data. This approach allows us to address three questions.

First, how many sources of variance underlie individual differences in the Recall Dynamics Functions, and what do these sources mean in terms of cognitive processes? Answering this question will place constraints on models of memory search (Underwood, 1975). Intuitively, one may predict that separate sources of variance contribute to each of the Recall Dynamics Functions (i.e., an initiation factor, a temporal contiguity factor, and a semantic contiguity factor). Under most models, however, each of the Functions results from multiple interacting mechanisms, making it unclear how many factors a model predicts. For example, under retrieved context models (e.g., Polyn, Norman, & Kahana, 2009), one parameter governs the influence of new temporal associations, another governs existing semantic associations, and both are scaled by a third parameter. In principle, individuals could differ on any or all of these parameters; thus, the model could predict that as few as one and as many as three sources of variance underlie the temporal and semantic Recall Dynamics Functions. The predictions of other models are similarly ambiguous (see the Discussion for more on this issue).

Second, do the Recall Dynamics Functions provide a complete description of the processes governing memory search? If so, they should contain all of the information needed to reconstruct an individual's overall probability of recall, which represents the outcome of all memory search processes. Answering this question will help identify gaps in our understanding of recall dynamics; if a substantial proportion of the variance in overall recall is unexplained, it suggests that the Recall Dynamics Functions miss important memory processes.

Finally, which memory processes are related to intelligence? Knowing which aspects of the Recall Dynamics Functions correlate with intelligence will allow modelers to test whether the corresponding parameter in a model is also correlated with intelligence. For the individual differences literature, knowing how the internal factor structure of recall dynamics relates to intelligence will advance the goal of understanding why memory predicts intelligence.

Methods

Participants

The data reported here are from the Penn Electrophysiology of Encoding and Retrieval Study (PEERS). PEERS aims to assemble a large database on the electrophysiological correlates of memory encoding and retrieval. The present analyses are based on the 141 college students (age 17–30) who had completed Experiment 1 of PEERS as of September 2013. Participants were recruited through a two–stage process. First, we recruited right-handed native English speakers for a single session to introduce participants to EEG recordings and the free recall task (EEG data are not reported here). Participants who completed this introductory session were invited to enroll in the full study, on the condition that they did not make an excess of eye movements during item presentation epochs of the experiment and their probability of recall was less than 0.8. Approximately half of the subjects recruited for the preliminary session qualified for, and agreed to participate in, the multi-session study. Participants were consented according the University of Pennsylvania's IRB protocol and were compensated for their participation.

PEERS Experiment 1

For completeness, we provide a full description of PEERS Experiment 1, but note that our primary analysis was conducted on the immediate free recall data. Participants performed a free recall experiment consisting of 1 practice session and 6 subsequent experimental sessions (the practice session is not included in the analyses reported below, though we note that including it produces almost identical results and does not change any conclusions). Each session consisted of 16 lists of 16 words presented one at a time on a computer screen. Each study list was followed by an immediate free recall test and each session ended with a recognition test. The practice session and half of the experimental sessions were randomly chosen to include a final free recall test before recognition, in which participants recalled words from any of the lists from the session.

Words were either presented concurrently with a task cue, indicating the judgment that the participant should make for that word, or with no encoding task. The two encoding tasks were a size judgment ("Will this item fit into a shoebox?") and an animacy judgment ("Does this word refer to something living or not living?"), and the current task was indicated by the color and typeface of the presented item. Using the results of a prior norming study, only words that were clear in meaning and that could be reliably judged in the size and animacy encoding tasks were included in the pool. There were three conditions: no-task lists (participants did not have to perform judgments with the presented items), single-task lists (all items were presented with the same task), and task-shift lists (items were presented with either task). The first two lists were task-shift lists, and each list started with a different task. The next fourteen lists contained four no-task lists, six single-task lists (three of each of the task), and four task-shift lists. List and task order were counterbalanced across sessions and participants.

Each word was drawn from a pool of 1638 words. Lists were constructed such that varying degrees of semantic relatedness occurred at both adjacent and distant serial positions. Semantic relatedness was determined using the Word Association Space (WAS) model described by Steyvers, Shiffrin, and Nelson (2004). WAS similarity values were used to group words into four similarity bins (high similarity: $\cos\theta$ between words > 0.7; medium–high similarity, $0.4 < \cos\theta < 0.7$; medium-low similarity, $0.14 < \cos\theta < 0.4$; low similarity, $\cos\theta < 0.14$). Two pairs of items from each of the four groups were arranged such that one pair occurred at adjacent serial positions and the other pair was separated by at least two other items.

For each list, there was a 1500 ms delay before the first word appeared on the screen. Each item was on the screen for 3000 ms, followed by jittered (i.e., variable) inter-stimulus interval of 800–1200 ms (uniform distribution). If the word was associated with a task, participants indicated their response via a keypress. After the last item in the list, there was a jittered delay of 1200–1400 ms, after which a tone sounded, a row of asterisks appeared, and the participant was given 75 seconds to attempt to recall aloud any of the just-presented items.

If a session was selected for final free recall, following the immediate free recall test from the last list, participants were shown an instruction screen for final free recall, telling them to recall all the items from the preceding lists. After a 5 s delay, a tone sounded and a row of asterisks appeared. Participants had 5 minutes to recall any item from the preceding lists.

After either final free recall or the last list's immediate recall test was a recognition test, which is not considered here (for full details see Lohnas & Kahana, in press).

PEERS Experiment 2

PEERS Experiment 2 was used to test the generalizability of our factor analysis. Of the 141 participants included in our main analyses, 127 also completed Experiment 2, which differed from Experiment 1 as described below. There was one practice session (not analyzed) followed by 6 experimental sessions each consisting of 12 study lists of 16 words. Experiment 2 included a mix of immediate recall lists, delayed recall lists (in which the final word was followed by a distractor task), and continual distractor lists (in which each word was followed by a distractor task). Distractor tasks consisted of answering math problems $A+B+C=?$, where A, B, and C were positive, single-digit integers, though the answer could have been one or two digits. When a math problem was presented on the screen, the participant typed the sum as quickly as possible. The task was self-paced, such that a participant may have been presented with, but not responded to, a problem at the end of the distraction interval. Participants were given a monetary bonus based on the speed and accuracy of their responses. In the first two trials, participants performed free recall with one trial having a distractor period following the last word presentation for 8 s. For the other of the first two trials, participants performed the distractor task for 8 s prior to and following each word presentation. In the remaining 10 trials, participants performed free recall with 5 possible time durations for the between-item and end-of-list distractor tasks. As listed here, the first number indicates the between-list distractor duration and the second number indicates the end-of-list distractor duration, both in seconds: 0-0 for immediate recall, 0-8 or 0-16 for delayed recall, and 8-8 or 16-16 for continual distractor recall. A 0 s distractor refers to the typical, non-filled duration intervals as described for Experiment 1. Within each session, 50% of the lists were randomly chosen to be task-switch lists, and the other half were single-task lists.

Intelligence Testing

The Wechsler Adult Intelligence Scale (WAIS) IV (Wechsler, 2008) was administered to 101 of the participants who completed Experiment 1. WAIS testing was conducted by a trained clinical psychologist in one–on–one sessions after completing all free recall sessions. We omitted the working memory index of the WAIS as we were concerned that participants' extensive practice with free recall would artificially inflate their scores.

Results

Identifying Sources of Variance in Recall Dynamics

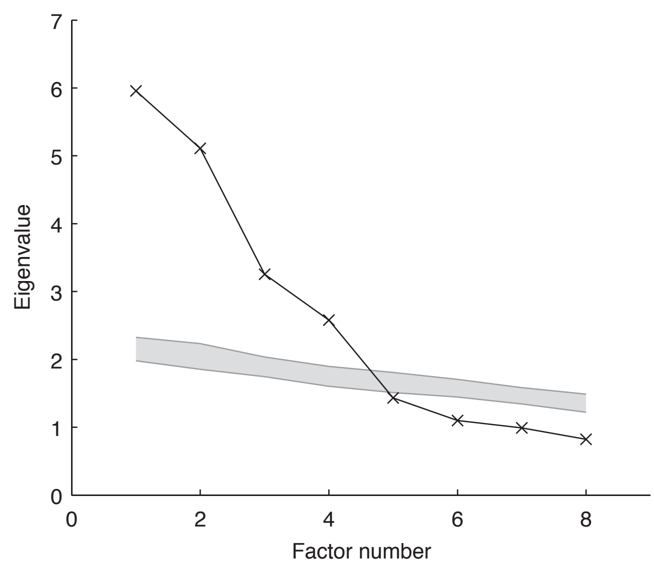

Figure 2. Results of Monte Carlo factor identification procedure (Glorfeld, 1995; Horn, 1965). The shaded region represents the middle 95% of the distribution of eigenvalues from 1000 simulated datasets that contained no factor structure. The line represents eigenvalues for the actual data. Only Factors 1–4 in the actual data fall above the shaded region indicating they explain more variability than expected by chance, and that the data contain 4 significant sources of variance.

The Recall Dynamics Functions are composed of 36 variables (16 points in the PFR, and 10 points each in the Lag-CRP and Semantic-CRP). Assume for a moment that only two cognitive processes contribute to individual differences

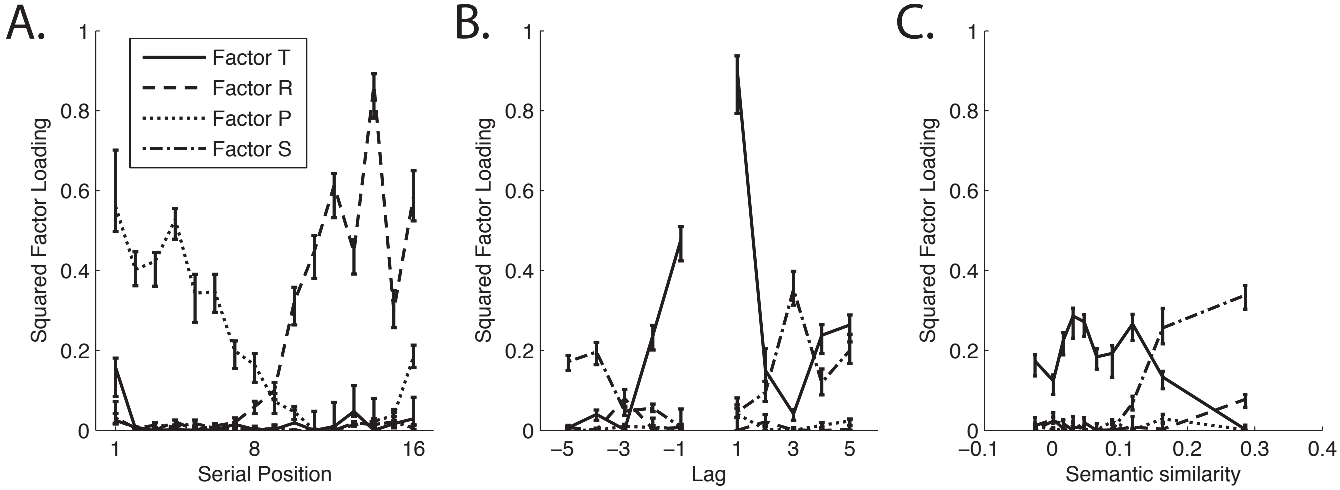

Figure 3. Squared loadings mapped onto the Recall Dynamics Functions: **A**. Probability First Recall curve. **B**. Lag-Conditional Response Probability curve. **C**. Semantic-Conditional Response Probability curve. The values at each point indicate the percentage of variance across participants that is accounted for by each factor. Error bars represent 99% confidence intervals (see text for details on how confidence intervals were computed). $N = 141$.

on these 36 variables. Perhaps there is a process that controls how participants initiate recall and another that controls how likely they are to make near–temporal transitions after initiation. If participants vary independently on these two processes, we should be able to describe an individual's recall dynamics functions with two sets of numbers. The first set of numbers would be a set of loadings that describe how much each processes contributes to each of the 36 variables (e.g., a process that controls recall initiation would likely have strong loadings for the PFR, but weaker loadings for the Lag-CRP and the Semantic-CRP). The second set of numbers needed to describe an individual's functions would be a score for each cognitive process that represents where in the distribution of variation on that process the individual lies (e.g., do they show strong or weak temporal contiguity).

Factor analysis allows us to extract these two sets of numbers from the data. Specifically, the data analyst specifies how many sources of variance, or factors, there are in the data. For each of these factors, the factor analysis algorithm determines a *factor loading* for each variable, which tell us how much that factor contributes to that variable. The algorithm also determines a *factor score* for each participant, which gives the participant's position in the distribution of variation on that factor. Combining the factor scores with the factor loadings allows us to recreate a participant's data from the equation:

$$\hat{\mathbf{x}} = \mu + \Lambda \mathbf{f} + \mathbf{e}, \tag{1}$$

Where $\hat{\mathbf{x}}$ is vector of predictions for the participant's 36 Recall Dynamics Function data points, μ is a vector of the means of the variables, Λ is a matrix of factor loadings with one row for each variable and one column for each factor, $\mathbf{f}$ is a vector of factor scores (one for each factor), and $\mathbf{e}$ is vector of error terms, which represent variation in the data that is not accounted for by the factors. The algorithm provides a single set of factor loadings and a set of factor scores for each participant by determining which set of factor loadings have the highest likelihood of having produced the actual data (i.e., a maximum likelihood estimate).

One limitation of factor analysis is that it requires the data analyst to specify how many factors to extract. Of course, we do not know how many processes actually contribute to the Dynamics Functions. If we extract 2 factors when there are really more than 2 factors, the recreated dataset will not capture all of the variability in the true dataset. More generally, if we call the true number of factors m, then we should be able to represent most of the variability in the Dynamics Functions using only m factors. The question then becomes, what is the value of m.

To find the appropriate value of m, we use a method (Glorfeld, 1995; Horn, 1965) that starts with the intuition that a dataset with m underlying sources of variance (i.e., factors) will look different than a dataset with uncorrelated variables. Each new factor you add to the factor analysis model will account some additional proportion of the variance in the original data. This proportion of variance is measured by the factor's eigenvalue, which represents how much of the overall variability in the data can be explained by that factor. If the data contain no true factors, then each factor should account for only a small proportion of the variance in the data (i.e., each factor should have a small eigenvalue). By contrast, if the data actually have m factors, then the first m factors should account for a considerable proportion of the variance (i.e., should have large eigenvalues), and factors > m should account for less variability. In other words, the first m eigenvalues of a dataset with m factors should be higher than the corresponding eigenvalues for uncorrelated datasets. Thus, to find the value of m we need to determine what the eigenvalues would be if there were no true factors, and then compare these with the eigenvalues obtained from the participants' data.

We can determine the expected eigenvalues for uncorrelated data by running a factor analysis on a simulated dataset that has the same means and variance as the actual data, but in which the variables are uncorrelated. Taking the PFR as an example, the value of a person's PFR at serial position 1 will be correlated with the value of their PFR at serial position 2. We want to create a random PFR that has the same shape, but lacks this correlation structure. To create such a simulated PFR curve we start with serial position 1 and draw values (one for each actual participant) from a random distribution with a mean and variance equal to the mean and variance of the actual PFR at serial position 1. We do the same for serial position 2, and so on. Because each serial position is drawn from an independent random distribution, there will be no correlation between serial positions. We can create simulated lag-CRPs and simulated semantic-CRPs in the same way, providing us with a full set of simulated Recall Dynamics Functions.

We can then run a factor analysis on the simulated dataset and save the eigenvalues. By repeating this procedure for 1000 simulated sets we build a distribution of expected eigenvalues for uncorrelated data. We then compare eigenvalues computed from the actual data with this distribution: If the data have m factors, the eigenvalues for the first m factors should lie above the 95th percentile of the simulated distribution but those for factors greater than m should not. Figure 2 indicates that 4 factors underlie the Recall Dynamics Functions.[1]

Linking Factors to Memory Processes

Next, we examined the factor loadings to link the factors to memory processes. The initial set of loadings returned by the factor analysis algorithm requires the factors to be orthogonal. However this initial solution can be "rotated" to make it more theoretically meaningful (Kline, 2005). To get a sense of how rotation works, imagine rotating the axis of a scatter plot so that the x-axis aligns with the regression line. Because there is no strong theoretical reason to believe that memory processes should be uncorrelated, we applied a oblique rotation (the Promax rotation) to the factor structure before examining the loadings. An oblique rotation allows factors to correlate. To ensure that the factor loadings are not biased by outliers and to provide confidence intervals on the loadings, we ran a jackknife procedure in which we ran the factor analysis multiple times, each time leaving one participant out of the sample (Clarkson, 1979). We used the distribution of loadings across samples to create 99% confidence intervals around the mean loadings. These mean loadings were then used to calculate factor scores using the Bartlett (1937) method.

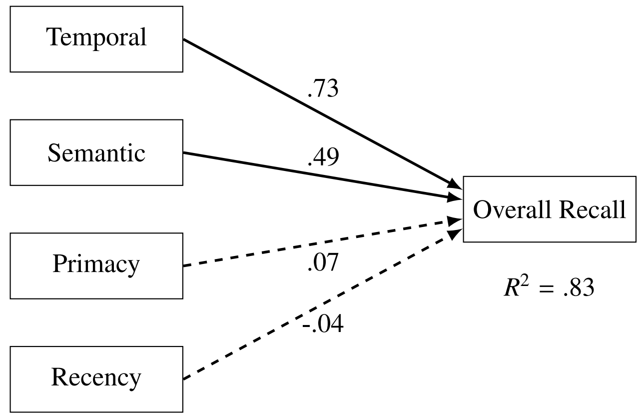

Figure 5. A path analysis model predicting overall recall from the four process factors. Each box represents a variable. The single-headed arrows connecting the variables represent hypothesized direct effects of one variable on another; the numbers next to single-headed arrows are standardized path coefficients and can be interpreted as standardized regression coefficients. The R^2 value is the proportion of variance in overall recall accounted for by the process factors. N = 141.

Each point in the Recall Dynamics Functions has one loading for each of the factors, and the square of these loadings tells us how much of the variance in that variable is explained by the factor (analogous to the R^2 in a regression). Figure 3 shows squared loadings mapped onto the original variables (Table 1 shows the non–squared loadings). The PFR loads primarily on factors 2 and 3, suggesting two major sources of individual differences in recall initiation. Early serial positions (positions less than 9) loaded onto Factor 3 suggesting that this factor measures the influence of primacy on recall initiation. Late serial positions (those greater than 9) loaded most strongly on Factor 2, suggesting that this factor measures the influence of recency on recall initiation. We therefore label these factors as the primacy and recency Process Factors. The finding that there are distinct primacy and recency factors is consistent with models that assume primacy and recency derive from separate mechanisms. For example, retrieved context models (e.g., Lohnas, Polyn, & Kahana, submitted; Polyn et al., 2009) assume primacy is due to increased attention to early list items whereas recency

[1]All of the factor analyses reported in this paper were conducted in MATLAB Release 2013a using the *factoran* function of the Statistics Toolbox. All path analyses reported here were conducted in R Version 3.0.1 using the lavaan package (Rosseel, 2012).

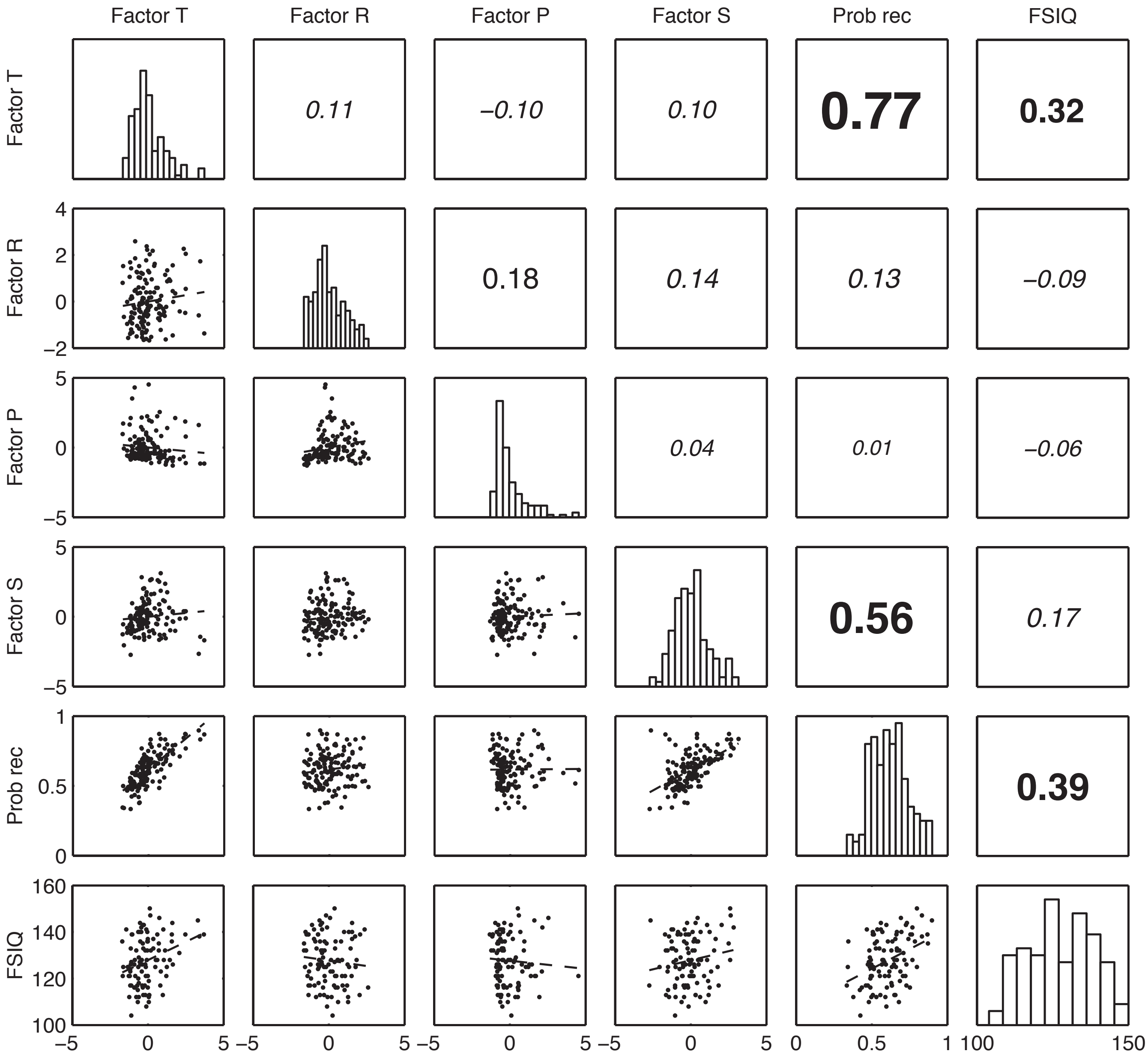

Figure 4. Correllelogram for the Process Factors (Factor T = Temporal Factor, Factor R = Recency Factor, Factor P = Primacy Factor, Factor S = Semantic Factor), overall recall, and WAIS-IV IQ. All correlations involving IQ have $N = 101$; all other correlations have $N = 141$. Plots on the diagonal show histograms of each variable; plots in the bottom triangle are scatter plots; the upper triangle shows the correlation coefficients. For correlation coefficients, the size of the typeface is proportional to the absolute size of the coefficient. Correlation values in *italics* are non–significant.

is due to mental context states at retrieval providing strong cues for recently presented items. We return to this point in the discussion. The pattern also fits with the finding that whereas most participants initiate from the end of the list, a minority tend to initiate from the beginning (Grenfell-Essam & Ward, 2012; Healey & Kahana, in press).

Post–initiation dynamics, as described by lag-CRP and semantic-CRP, load primarily on factors 1 and 4. Factor 1 is strongly related to near temporal transitions and far semantic transitions. By contrast, factor 4 loads most strongly on near semantic transitions and distant temporal transitions. Therefore we label factor 1 and factor 4 as the temporal and the semantic Process Factors respectively. The fact that the temporal factor has high loadings not just for near temporal

Table 1
Promax rotated factor loadings.

	Factor T	Factor R	Factor P	Factor S
Serial position		PFR		
1	0.40	−0.16	0.75	−0.18
2	0.07	−0.09	0.63	0.08
3	−0.01	0.02	0.65	0.11
4	0.05	−0.15	0.73	0.10
5	0.03	0.04	0.59	0.13
6	−0.03	0.11	0.59	0.08
7	−0.12	0.14	0.45	−0.06
8	−0.05	0.24	0.41	−0.06
9	−0.05	0.32	0.27	0.07
10	−0.14	0.57	0.22	0.00
11	−0.01	0.67	0.03	−0.01
12	−0.10	0.78	−0.00	0.02
13	0.22	0.67	−0.12	−0.14
14	−0.01	0.93	−0.15	−0.12
15	−0.13	0.55	−0.19	0.15
16	−0.17	−0.77	−0.43	0.08
Lag		Lag-CRP		
-5	−0.08	−0.09	−0.02	0.41
-4	−0.20	0.03	0.02	0.44
-3	−0.05	0.28	−0.10	0.22
-2	0.49	0.06	−0.09	0.24
-1	0.69	−0.09	−0.07	−0.09
1	0.95	−0.01	0.20	−0.22
2	0.39	0.15	−0.05	0.31
3	−0.20	0.04	−0.00	0.59
4	−0.49	−0.01	−0.11	0.35
5	−0.51	−0.04	0.15	0.45
Similarity Bin		Sem-CRP		
1	0.42	0.11	0.10	−0.05
2	0.35	0.17	0.04	−0.15
3	0.47	0.07	0.03	−0.13
4	0.54	0.08	−0.11	−0.02
5	0.52	0.09	−0.06	0.14
6	0.43	−0.02	−0.04	0.02
7	0.44	0.04	0.15	−0.04
8	0.52	−0.09	0.03	0.26
9	0.37	−0.06	0.17	0.51
10	0.05	−0.28	0.04	0.58

T = Temporal; R = Recency; P = Primacy; S = Semantic. Factors are order in decending order of eigenvalue

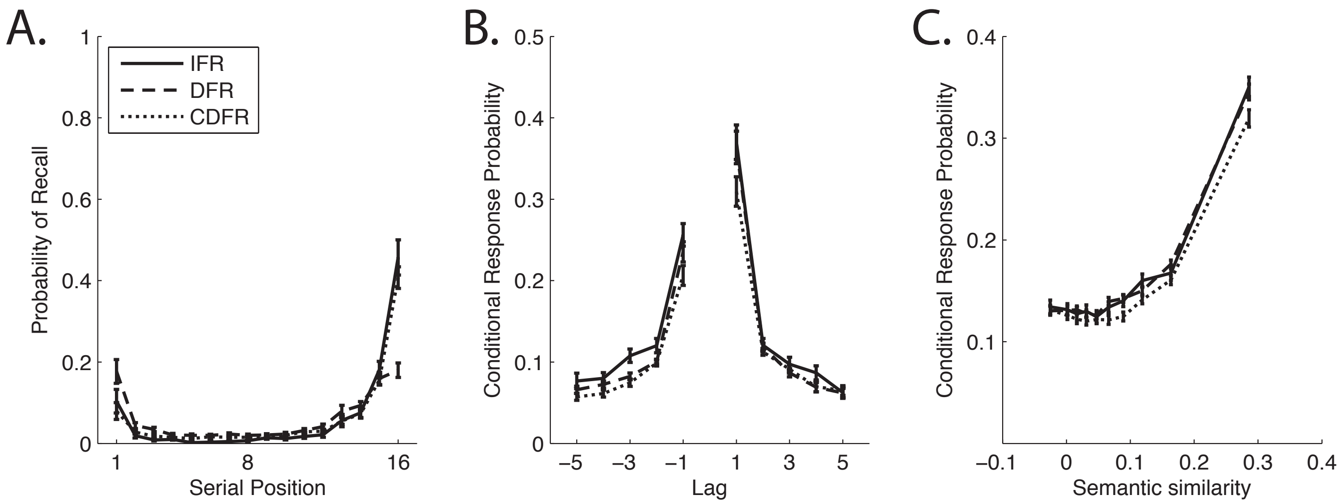

Figure 6. The Recall Dynamics Functions from the 127 participants who completed Experiment 2 of the Penn Electrophysiology of Encoding and Retrieval Study. **A**. Probability First Recall curve. **B**. Lag-Conditional Response Probability curve. **C**. Semantic-Conditional Response Probability curve. Within each panel the three lines correspond to Immediate Free Recall (IFR), Delayed Free Recall (DFR), and Continual Distractor Free Recall (CDFR). See the text for details on how these curves are created. Error bars are 95% within-subject confidence intervals (Loftus & Masson, 1994).

transitions but also for far semantic transitions (and that the semantic factor has high loading not just for near semantic transitions but also for distant temporal transitions), likely reflects a natural tradeoff between temporal and semantic clustering. Take the extreme example of a participant who shows perfect temporal clustering (i.e., recalls in perfect serial order) and makes no use of semantic associations. Because our lists are arranged so that most words with strong semantic associations are not temporally contiguous, such a participant would tend to make few transitions between close semantic associates which necessarily means they will show stronger contiguity for items with low semantic similarity than for items with strong semantic similarity.

For each participant we can use the factor loadings to derive a score for each of the factors, which can be used as that participant's measures of the primacy, recency, temporal, and semantic processes. Before we use the scores as individual differences measures, however, we must ensure they are reliable. To assess reliability we used the split–half technique. Specifically we used the factor loadings from the full dataset to compute a set of factor scores for half of the 6 sessions and another set of factor scores for the other half of the six sessions and correlated the two sets of scores. We repeated this procedure for 100 split–half samples to compute average split-half reliability. These reliabilities for the Primacy, Recency, Temporal, and Semantic factors (± a 95% confidence interval computed across the samples) were .86 ± .005, .88 ± .005, .82 ± .007, and .68 ± .006, respectively. The reliability of overall recall, computed using the same procedure, was .90 ± .005. These values are well above the common threshold of .6 for acceptable reliability. To determine how the 4 factors relate to each other, we correlated participants' scores across the factors (Figure 4). Despite our use of an oblique factor rotation, the factor scores remain largely uncorrelated. The lack of a correlation between temporal and semantic factors (see also Sederberg et al., 2010) suggests that temporal and semantic clustering arise from two distinct cognitive processes.

Using recall dynamics to predict overall accuracy

Overall accuracy reflects the outcome of all the processes that contribute to memory search. If the Recall Dynamics Functions provide a complete description of recall dynamics, we should be able to use the factors derived above to predict overall recall. Because none of these Recall Dynamics Functions reflect recall success directly, there is no a priori reason for the factors to correlate with recall success (Sederberg et al., 2010).

As shown in Figure 4, the temporal and semantic factors are both predictive of overall recall success, but the initiation factors are not. In a meta-analysis, Sederberg et al. (2010) found that for random word lists a temporal clustering summary score was moderately correlated with recall success, but that a semantic summary score was not, whereas in lists that include pairs of semantically related words, as did our lists, semantic clustering was predictive of recall success. This pattern of results suggests the relationship between clustering and recall success depends on the content of the lists. As we elaborate in the discussion, participants may dynamically tune their memory systems to up–weight associations that facilitate performance and down–weight those that do not.

To determine how well the factors account for overall recall accuracy, we can use all four factors to simultaneously predict recall. Figure 5 shows the results. The figure is essentially a simultaneous regression but we present it as a path analysis model to maintain consistency with more complex analyses we present later. Each box in the figure represents a variable, and the arrows connecting the boxes represent the influence of variables on each other, with the direction of the arrow giving the presumed direction of the effect (e.g., we assume that the process factors cause variance in overall recall). The numbers next to the paths are analogous to standardized beta weights in a regression. The figure shows that together the four factors account for 83% of the variability in overall recall, suggesting that the Recall Dynamics Functions provide a near–complete description of recall dynamics. No doubt other factors (e.g., idiosyncrasies in semantic relationships not captured by LSA) account for some proportion of overall recall, but apparently not more than 17%. Examining the individual paths, we see that the temporal and semantic factors were significant predictors of overall recall (solid lines represent significant paths), but that the primacy and recency factors were not.

Validating the factor structure

A strong test of the validity of the factor structure would be to use the factors computed above to predict performance on a second dataset. Of the 141 participants who completed Experiment 1 of PEERS, 127 also completed Experiment 2. We can use the data from this second experiment to validate the factor structure we discovered in the Experiment 1 data. One approach would be to independently rerun the entire factor analysis on the Experiment 2 data. A more stringent test, however, would be to use the factors derived in Experiment 1 to predict Experiment 2 performance.

Figure 6 shows the Recall Dynamics Functions for Experiment 2. Figure 7 shows the squared correlations between each Experiment 1 factor and the Experiment 2 Recall Dynamics Functions, which can be interpreted in the same way as the squared factor loadings in Figure 3. Examining Figure 7 reveals that the loading patterns are quite similar across the three versions of free recall in Experiment 2, and also quite consistent with the loading pattern observed in Experiment 1. The most notable deviation is that the primacy and recency factors explain less of the variability in the PFRs from Experiment 2 than they did for Experiment 1. Why would the recall initiation factors show less generalization across experiments than the post–initiation transition factors? We have recently shown that whereas all participants show qualitatively similar lag– and semantic–CRP functions, there is more variability in the shape of PFR functions, with most participants showing a recency heavy function but a minority showing either a non–monotonic "clustering" pattern or a primacy pattern (Healey & Kahana, in press). We suggested that the variability in PFR functions may reflect, in part, differences in strategy. It is possible that because Experiment 2 intermixes immediate recall trials, which tend to show a recency initiation pattern, with delayed trials which tend to show a recency initiation pattern (compare the PFR curves in Figure 6, participants may adopt different strategies when they begin Experiment 2, reducing the predictive power of the initiation factors derived from Experiment 1. In contrast, the fact that the temporal and semantic factors tend to correlate with the same Recall Dynamics Function points across all three recall tasks suggests that the same processes govern post–initiation dynamics regardless of distractor condition.

This interpretation of the relative consistencies of the factors across task type is supported by examining the ability of the Process Factors derived from Experiment 1 to predict overall recall on the three Experiment 2 tasks. Figure 8 shows path diagrams of these predictions. For immediate free recall (Figure 8A), the pattern is very similar to that seen in Experiment 1 (Figure 5), with the temporal and semantic factor both being significant predictors but the primacy and recency factors being non-significant. This strengthens our claim that variation in post recall dynamics is more diagnostic of episodic memory ability than variation in recall initiation patterns. An even stronger test is to use the factors from immediate free recall to predict delayed and continual distractor tasks. If the three tasks rely on the same mechanisms, as predicted by retrieved context models (e.g., Polyn et al., 2009), we would expect the quality of prediction to be quite high. If, however, the tasks rely on different processes (e.g., short– versus long–term memory; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005) the quality of prediction should suffer. Panels B and C of Figure 8 show that the quality of prediction is in fact, excellent: The immediate free recall factors predict 62% of the variability on delayed free recall, and 51% of the variability on continual distractor free recall. Once again, the temporal and semantic factors were significant predictors, but the primacy and recency factors were not.

The ability of the factor structure derived from IFR in Experiment 1 to predict delayed recall in Experiment 2 is particularly striking in that it suggests that any sources of variance uniquely related to a short–term buffer, which should be emptied by the distractor, account for a relatively small proportion of the variance in both overall recall and the dynamics functions. We note that the R^2 for continual distractor free recall is somewhat lower than for the other two conditions. This suggests the possibility that continual distractor recall may capture a source of individual differences that is not (as fully) captured by immediate free recall. One possibility is that participants may vary in the extent to which

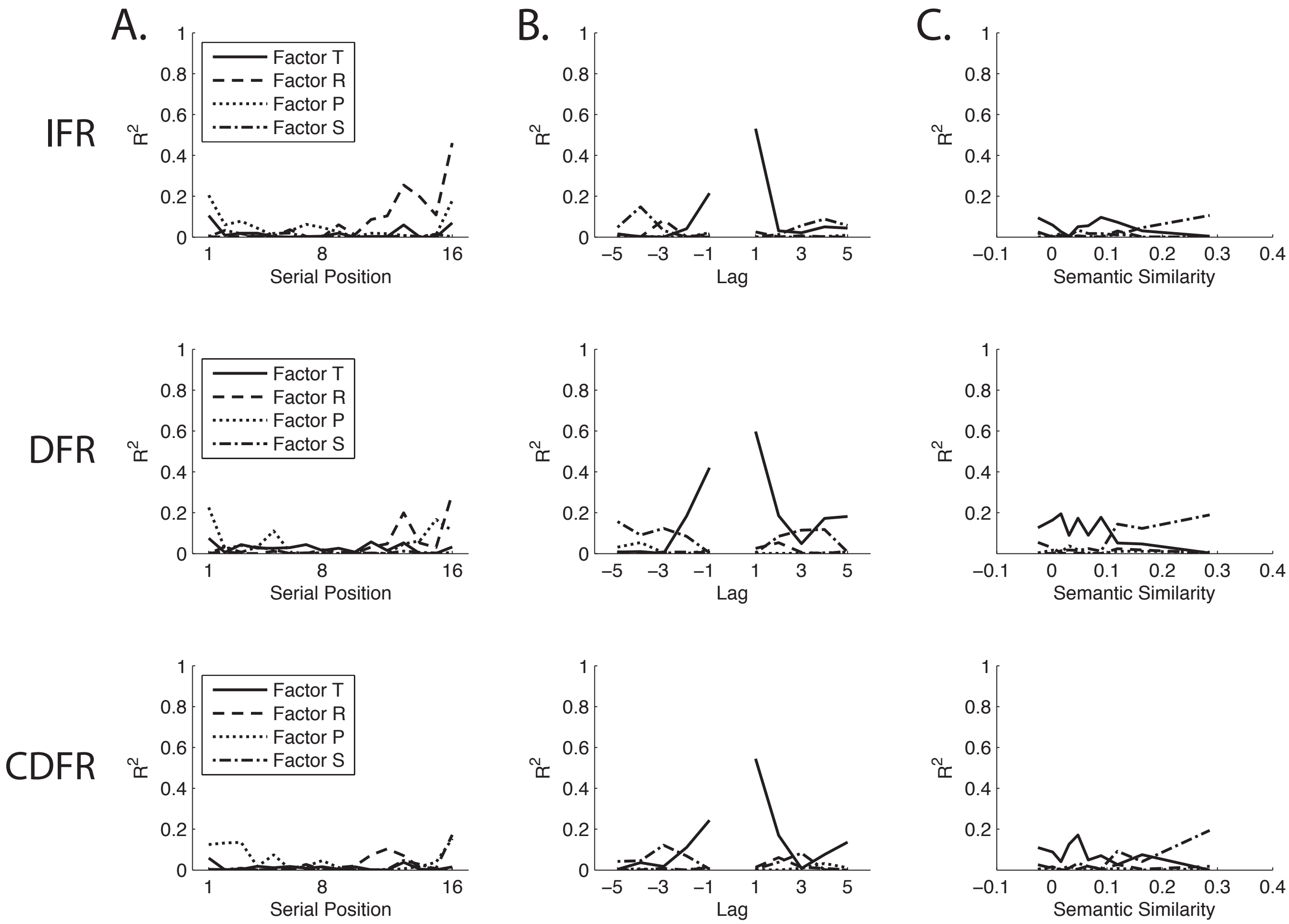

Figure 7. Squared loadings mapped onto the Recall Dynamics Functions of Experiment 2 of the Penn Electrophysiology of Encoding and Retrieval Study: **A**. Probability First Recall curve. **B**. Lag-Conditional Response Probability curve. **C**. Semantic-Conditional Response Probability curve. The first row shows Immediate Free Recall (IFR), the second row shows Delayed Free Recall (DFR), and the third row shows Continual Distractor Free Recall (CDFR). The values at each point indicate the percentage of variance across participants that is accounted for by each factor. $N = 127$

mental context drifts during distractor intervals.

Using recall dynamics to illuminate the memory/intelligence correlation

The distribution of IQ scores in our sample (Figure 4), while roughly normal, is above the population average. Despite this abbreviated range, which will tend to produce underestimates of the true correlations between memory and IQ, we found a correlation between overall recall and IQ of .39 (see Figure 4) within the range reported in meta–analyses (Ackerman et al., 2005). Translating this correlation into proportion of variance, overall recall accounted for 15% of the variability in IQ (Figure 9A). One of the main goals in the study of individual differences in memory is to determine which memory processes are responsible for the variance shared between overall recall and IQ. That is, *why* does overall recall success predict IQ. Do the Process Factors we have identified here help answer this question?

The first step is to examine the correlations between the factors and IQ. Figure 4 shows that both the temporal and the semantic factor are significantly correlated with IQ but that neither the primacy nor the recency factors are. Next we ran a simultaneous regression using the four factors to predict IQ. Together, the factors accounted for 14% of the variability in IQ scores (Figure 9B), with the temporal factor being the strongest predictor. Note that overall recall and the process factors account for almost the identical proportions of variance in IQ (i.e., 15% and 14% respectively). If the processes factors are measuring individual differences in the memory processes that drive the correlation between over-

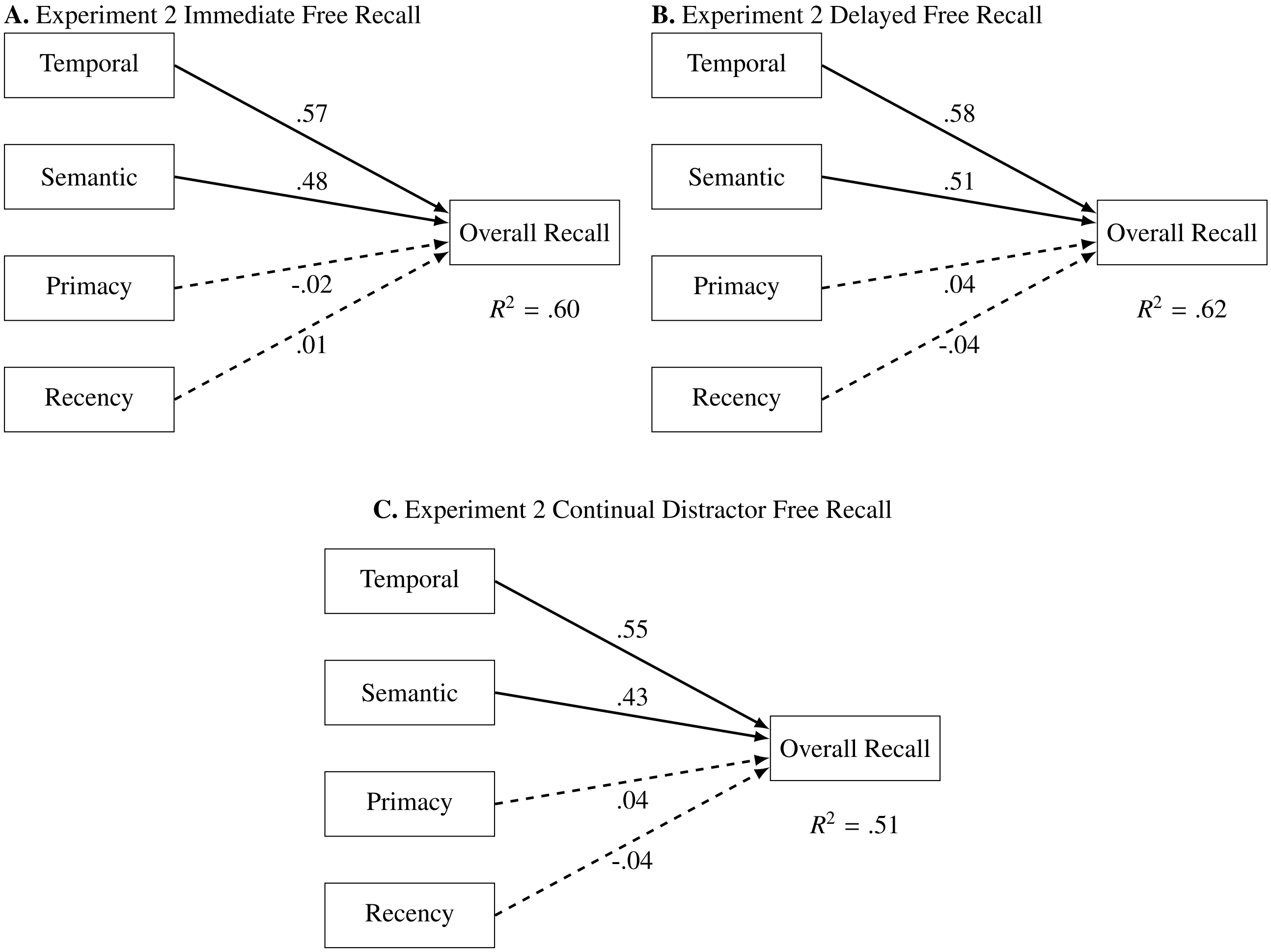

Figure 8. Path analysis models predicting overall recall from the four process factors for each condition of Experiment 2: **A.** Immediate Free Recall, **B.** Delayed Free Recall, and **C.** Continual Distractor Free Recall . Each box represents a variable. The single-headed arrows connecting the variables represent hypothesized direct effects of one variable on another; the numbers next to single-headed arrows are standardized path coefficients and can be interpreted as standardized regression coefficients. The R^2 values are the proportion of variance in overall recall accounted for by the process factors. N = 127.

all recall and IQ, we should find that the factors and overall recall account for overlapping portions of the variability in IQ.

To test our if our factors fully account for the relationship between memory and IQ we conducted a commonality analysis (Nimon, Lewis, Kane, & Haynes, 2008). A commonality analysis takes the total variance in one variable explained by a set of predictor variables and attempts to break it down into variance that is uniquely accounted for by one predictor (but not others), and variance that is explained by several predictors (i.e., shared variance). Commonality analysis has previously been used to partition the variance in IQ explained by working memory versus episodic memory tasks (Unsworth & Spillers, 2010). If you imagine all of the variation in IQ as a pie, the variability accounted for by overall recall would be a slice (15%) of the pie. The bar in Figure 10 represents that slice of the pie. We can further divide the slice into parts that represent variability uniquely explained by recall (but not the factors) and variance that is common to overall recall and the factors. Our prediction is that the portion unique to overall recall will be small. Consistent with our prediction, less than 20% (i.e., less than 3% of the entire IQ pie) of the variance was unique to overall recall. That is, overall recall and the factors account for almost completely overlapping variance in IQ suggesting that recall dynamics capture the processes that allow memory to predict IQ.

This view of the relationship among the factors, overall recall and IQ is made explicit in the path analysis model in Figure 11. As we discussed in the introduction the standard approach to determining whether a particular cognitive process accounts for the relationship between memory and IQ is to statistically control for variation in a third task that mea-

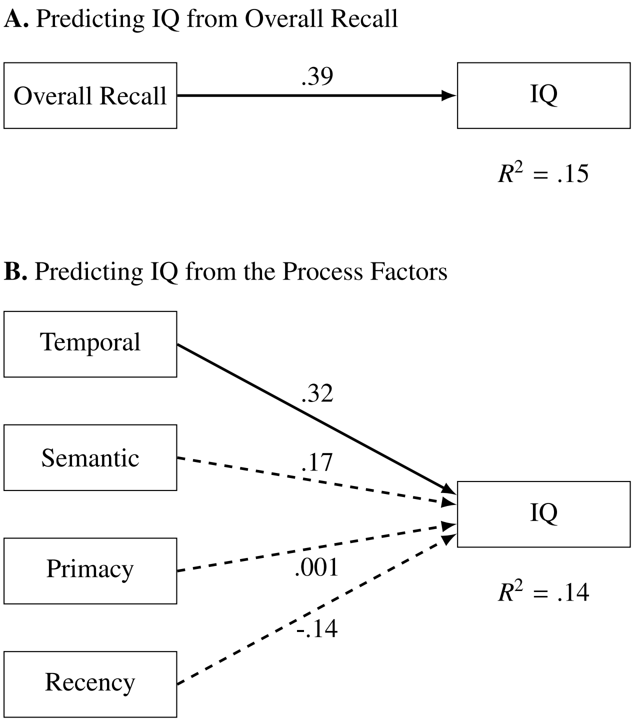

Figure 9. Path analysis model predicting IQ from Overall Recall (**A**) and the four process factors (**B**). Each box represents a variable. The single-headed arrows connecting the variables represent hypothesized direct effects of one variable on another; the numbers next to single-headed arrows are standardized path coefficients and can be interpreted as standardized regression coefficients. The R^2 values are the proportion of variance in overall recall accounted for by the process factors. N = 101.

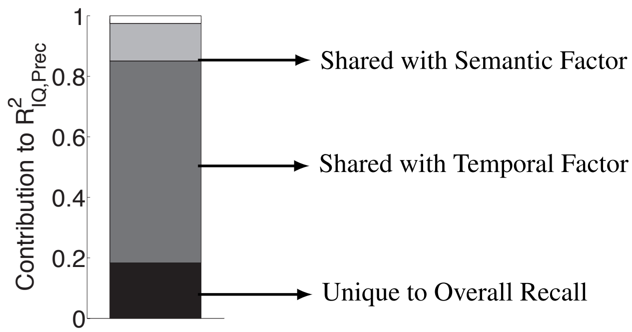

Figure 10. Commonality analysis of correlation between overall recall and IQ. The bar represents all of the variability in IQ that is accounted for by overall recall (i.e., 15%). The shaded regions represent the portion of that variability that is uniquely accounted for by overall recall, and the portion that is shared with the process factors. There is a small unique contribution of overall recall, but most is shared with the temporal and semantic factors. A small portion (the white region at the top of the bar) is shared with the primacy or recency factors or with multiple factors. N = 101.

sures the process in question. This mediation logic assumes that the third task provides a purer, or more sensitive, measure of the process in question than does either memory or IQ. By contrast we have extracted measures of memory processes directly from detailed measures of task performance rather than using a non–memory proxy task. That is, overall recall contains the same information as our process factors but compresses the information into a single measures; our factors uncompress the data.

Consistent with the logic that the factors represent the processes contributing to overall performance, the path model in Figure 11 includes direct paths (i.e., single-headed directly connecting boxes) from each factor to overall recall, which in turn has a direct path to IQ. Each factor also has an indirect path to IQ via their influence on overall recall. In the language of path analysis, our prediction is that these indirect effects of the process factors on IQ via overall recall should account for the bulk of the direct effect of overall recall on IQ. As expected, the direct effects of the factors on IQ are all non-significant. Indeed, constraining the direct paths from each of the factors to IQ to be equal to zero, provided very a very good fit to the data ($\chi^2(4) = 2.3, p = 0.68$). That is, none of the factors directly contribute to variation in IQ. Instead their effects are mediated by their influence on overall recall.

Discussion

Performance on memory tasks predicts many other cognitive abilities (Daneman & Carpenter, 1980; Jacobs, 1887; Mogle et al., 2008). To better understand the role of retrieval processes in producing these correlations we examined individual differences in measures of recall dynamics. We found that four distinct factors contribute to individual differences in memory search: a tendency to initiate recall with end–of–list (recency) items, a tendency to initiate with early–list (primacy) items, a temporal factor corresponding to transitions mediated by temporal associations, and a semantic factor corresponding to transitions mediated by long–standing semantic associations. We show that the four factors account for 83% of the variability in overall recall, suggesting that they provide a near–complete description of the processes that contribute to individual differences in recall success. To validate this factor structure we showed that the factors computed from immediate free recall in PEERS Experiment 1 account for 60%, 62%, and 51% of the variance in immediate, delayed, and continual distractor recall, respectively, in PEERS Experiment 2. Moreover, the factors accounted for over 80% of the relationship between memory and IQ, with the temporal factor being the most important single factor.

Implications for Models and Theories

The factor structure reported here places a new class of constraints on memory models. The dominant approach to model validation has been to fit models to data averaged across participants. An accurate model of memory search, however, should also account for differences among individuals: The covariance of the data. To our knowledge, no attempt has been made to fit the covariance structure of recall dynamics. Some preliminary observations are possible, however. For example, under dual–store models, temporal contiguity effects emerge from items spending time together in short-term memory; short-term memory also powerfully influences the tendency to initiate recall from the end of list, which may suggest temporal contiguity and initiation would share variance. Under retrieved context models, there is also reason to predict contiguity and initiation should be correlated, as both mechanisms are influenced by a common context drift rate parameter. Contrary to these intuitive predictions, we found that recall initiation and temporal contiguity are largely independent. Future modeling work should explore whether existing models can simulate the factor structure discovered here.

For the individual difference literature, our results provide a fresh perspective on which memory processes correlate with IQ. Most of the correlation between memory and IQ is accounted for by individual differences in temporal contiguity: Individuals who show stronger clustering tend to have higher IQs. This finding poses a challenge to theories that assume limitations in working memory are key. Under these theories, higher–IQ individuals should have more capacity to hold items in memory, either due to larger buffers (or "foci of attention"; Cowan, 2010; Oberauer, 2002) or to more efficient attentional processes (Hasher et al., 2007; Kane et al., 2007). Under most dual–store memory theories, temporal associations are formed between items that spend time together in working memory. An intuitive prediction of this view is that individuals who hold more items in working memory at a time will form longer–range temporal associations. Longer–range temporal associations would lead to more transitions at longer lags (i.e., a more shallow lag-CRP curve) and fewer transitions at the short lags that load strongly on the temporal factor. Therefore, one would expect individuals with large working memory capacities to have lower temporal factor scores. This prediction contradicts the positive correlation we found between temporal clustering and overall recall (see also Spillers & Unsworth, 2011). The ability of the factors derived from immediate free recall in Experiment 1 to account for roughly equal amounts of variance in both immediate and delayed recall in Experiment 2 (i.e., 60% and 62% respectively) is also informative on this point, as from a dual–store perspective, one may have predicted that variance related to the short–term buffer would be more important in predicting performance in immediate than delayed recall.

Temporal Contiguity and Intelligence

The temporal factor, corresponding to a tendency to make near temporal transitions, was most predictive of memory accuracy and IQ. What does the importance of temporal contiguity tell us about memory and IQ? A retrieved context model of the contiguity effect suggests a novel theory. To illustrate we must first describe how temporal contiguity arises in such models: When an item is presented it becomes associated with an internal context representation. The context representation changes as items are presented, but in an autocorrelated fashion so that items presented in temporal proximity become associated with similar contextual states. During retrieval, context is used as a cue. When item i is recalled, its associated context is retrieved and integrated into the context representation that cues the next item. This retrieved context is a strong cue for items presented near i as those items were associated with similar contexts, giving rise to the temporal contiguity effect.

Recall success may depend on a participant's ability regulate the drift of their context representation such that each recalled item retrieves a context that serves as an effective cue for another list item. Such control may be exercised by tuning the relative influence of different types of associations (Healey & Kahana, in press). For lists of completely unrelated words, it would be optimal to have temporal associations dominate contextual drift. Consistent with this suggestion, Sederberg et al. (2010) found that temporal but not semantic clustering correlated with recall for random lists. In lists that contain moderate semantic associations, such as those used here, allowing semantics to have a moderate influence on contextual drift would make sense, consistent with our finding that the semantic factor was correlated with recall ($r = .56$). Such tuning likely occurs outside of conscious awareness in response to experience with the task. For example, Golomb, Peelle, Addis, Kahana, and Wingfield (2008) found that across multiple learning trials in a serial recall task, young adults showed progressively stronger temporal clustering and progressively weaker semantic clustering. That is, with experience participants came to rely on those types of associations that facilitated performance. The converse of guiding context to favorable states is preventing it from drifting to states that make non-list items *more* accessible. Such "context gating" could be used to demarcate list boundaries, making task-irrelevant memories inaccessible, and may reflect the computational basis of the ability to resolve interference (Healey, Campbell, Hasher, & Ossher, 2010).

How do context regulation and gating relate to intelligence? A simple answer is that any complex task requires selective memory access. A deeper answer, however, is sug-

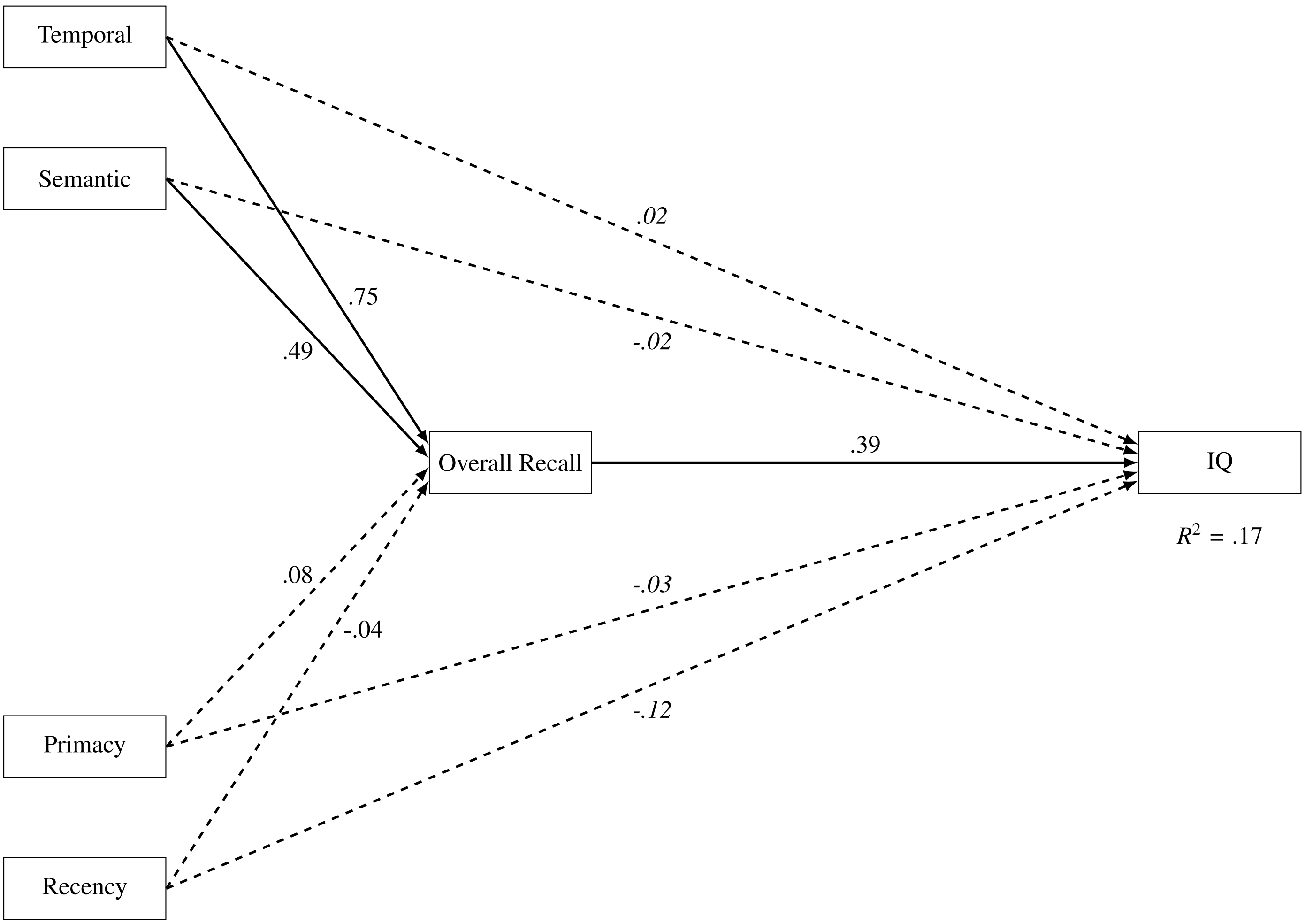

Figure 11. Path analysis model of the effects of the process factors and overall recall on IQ. Each box represents a variable. The single-headed arrows connecting the variables represent hypothesized direct effects of one variable on another; the numbers next to single-headed arrows are standardized path coefficients and can be interpreted as standardized regression coefficients. The direct paths from the factors to IQ were constrained to zero in the final model, the estimates from the non–constrained model are provided in *italics* for completeness. The R^2 values are the proportion of variance in Overall Recall accounted for by the process factors. N = 101.

gested by Duncan's (2010) idea of a multiple demand system. Under this view, frontal neurons instantiate distinct connection patterns reflecting current task demands, with orthogonal connection patterns across different phases of an experiment forming boundaries between different tasks. This notion of rapidly changing frontal networks adapting to task demands is closely related to the notion of an internal context representation that dynamically modulates the accessibilities of various memories. It may be that the ability to use internal contextual representations to dynamically gate access to relevant memories and response tendencies is a basic computational principle across tasks and a key component of intelligence.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60.

Baddeley, A. (2003). Working memory: looking back and forward. *Nature Reviews Neuroscience*, *4*(10), 829–839.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, *28*, 97–104.

Beier, M. E., & Ackerman, P. L. (2004, January). A reappraisal of the relationship between span memory and intelligence via "best evidence synthesis". *Intelligence*, *32*(6), 607–619.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229-240.

Carlson, H. B. (1937). Factor analysis of memory ability. *Journal Of Experimental Psychology*, *21*(5), 477–492.

Carroll, J. D. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Christal, R. E. (1958). Factor analytic study of visual memory. *Psychological Monographs: General and Applied*, *72*, 1-24.

Clarkson, D. (1979). Estimating the standard errors of rotated factor loadings by jackknifing. *Psychometrika*, *44*(3), 297-314.

Conway, A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. (2007). *Variation in working memory*. Oxford University Press.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *The American Psychologist*, *12*, 671–684.

Daneman, M., & Carpenter, P. A. (1980). Individual difference in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*(4), 422–433.

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

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, *54*, 180-187.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends In Cognitive Sciences*, *14*(4), 172–179.

Games, P. A. (1962). A factorial analysis of verbal learning tasks. *Journal Of Experimental Psychology*, *63*(1), 1–11.

Glorfeld, L. W. (1995). An improvement on Horn's Parallel Analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*, 377–393.

Golomb, J. D., Peelle, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition*, *36*(5), 947–956.

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal Of Memory And Language*, *67*(1), 106–148.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York: Oxford University Press.

Healey, M. K., Campbell, K. L., Hasher, L., & Ossher, L. (2010, October). Direct evidence for the role of inhibition in resolving interference in memory. *Psychological science*, *21*(10), 1464–1470.

Healey, M. K., & Kahana, M. J. (in press). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*.

Healey, M. K., & Miyake, A. (2009). The role of attention during retrieval in working-memory span: a dual-task study. *Quarterly journal of experimental psychology (2006)*, *62*(4), 733–745.

Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition*, *3*, 197-209.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.

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). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*, 85–98.

Jacobs, J. (1887). Experiments on "prehension". *Mind*, *os-12*(45), 75–79.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*, 103–109.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. Towse (Eds.), (pp. 21–48). New York: Oxford University Press.

Kane, M. J., & Miyake, T. M. (2008). Individual differences in episodic memory. In J. Byrne (Ed.), *Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference.* (pp. 773–785). Boston: Elsevier.

Kelley, H. P. (1964). Memory abilities: A factor analysis. *Psychometric Monographs*, *11*.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, *34*, 419-426.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.

Lohnas, L. J., & Kahana, M. J. (in press). Parametric effects of word frequency effect in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (submitted). Expanding the scope of memory search: Intralist and interlist effects in free recall.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, England: Cambridge University Press.

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological Science*, *19*, 1071-1077.

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*(2), 457–466.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421.

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

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, *140*(3), 464–487.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, *4*, 28-34.

Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, *48*.

Sederberg, P. B., Miller, J. F., Howard, W. H., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, *38*(6), 689–699.

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal Experimental Psychology: Learning, Memory and Cognition*, *37*(6), 1532–1539.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer.* Washington, DC: American Psychological Association.

Towse, J., Hitch, G., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal Of Memory And Language*, *39*(2), 195–217.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal Of Memory And Language*, *28*(2), 127–154.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *The American psychologist*, *30*(2), 128–134.

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General*, *107*, 393-419.

Unsworth, N. (2009). Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall. *Memory and Cognition*, *37*(6), 837–849.

Unsworth, N., Brewer, G., & Spillers, G. (2011). Inter- and intraindividual variation in immediate free recall: An examination of serial position functions and recall initiation strategies. *Memory*, *19*(1), 67–82.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity–fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin and Review*, *16*(5), 931–937.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: the role of controlled search. *Memory and Cognition*, *41*(2), 242–254.

Unsworth, N., & Spillers, G. J. (2010, May). Working memory capacity: Attention control, secondary memory, or both? a direct test of the dual-component model. *Journal Of Memory And Language*, *62*(4), 392–406.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson.