A Source Activation Theory of Working Memory: Cross-task Prediction of Performance in ACT-R

Marsha C. Lovett **Larry Z. Daily** **Lynne M. Reder**
Carnegie Mellon University

Abstract

Over the decades, computational models of human cognition have advanced from programs that produce output similar to that of human problem solvers to systems that mimic both the products and processes of human performance. In this paper, we describe a model that achieves the next step in this progression: predicting individual participants' performance across multiple tasks after estimating a single individual difference parameter. We demonstrate this capability in the context of a model of working memory, where the individual difference parameter for each participant represents working memory capacity. Specifically, our model is able to make zero-parameter predictions of individual participants' performance on a second task after separately fitting performance on a preliminary task. We argue that this level of predictive ability offers an important test of the theory underlying our model.

This work was supported by Air Force Office of Scientific Research Grant F49620-97-1-0455 and by Office of Naval Research Grant N00014-95-0233. We would like to thank Scott Filippino and Becky Stoltzer for their assistance in collecting the data and Christian Lebiere for many useful discussions. We would also like to thank John R. Anderson for comments on an earlier draft of this paper. Correspondence concerning this article should be addressed to Marsha C. Lovett, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent via the Internet to Lovett@cmu.edu.

A Source Activation Theory of Working Memory: Cross-task Prediction of Performance in ACT-R

Over the decades, computational models of human cognition have advanced from programs that could produce output similar to that of human problem solvers to systems that mimic both the products and processes of human performance. Indeed, the fidelity of recent computational models has increased to the point where models can reproduce multiple dependent measures collected across several variations of a task. These models have matched both aggregate behaviors and individual differences within a given task. In this paper, we describe a model that achieves the next step in this progression: predicting individual participants' performance across multiple tasks after estimating a single individual difference parameter. We demonstrate this capability in the context of a model of working memory, where the individual difference parameter for each participant represents working memory capacity. Specifically, our model is able to make zero-parameter predictions of individual participants' performance on a second task after separately fitting performance on a preliminary task. This level of predictive ability has not been attained by other computational models, and yet we argue that it offers an important test of the theory underlying any computational model.

The paper is organized as follows. In the next section, we present a brief historical sketch of how computational models have attained better and better fits to observed data while accounting for more and more features of the data. We then provide an overview of important working memory phenomena, highlighting three requirements that a model of working memory must exhibit in order to fully capture these phenomena. One of these requirements involves representing an individual's working memory capacity in such a way that the representation captures the similarities in that individual's performance across multiple tasks. Next, we present the details of our working memory model. Finally, we describe several experiments we conducted to test our model, and we show that the model meets the most stringent of our stated requirements.

Progress in Computational Modeling

Since the early demonstrations that computers could simulate intelligent behavior, there has been enormous progress in the computational modeling of human cognition. Some of the earliest work in this area, completed in the 1960's, is presented in the volume *Computers and Thought*, edited by Feigenbaum and Feldman (1963). There, papers like that by Newell, Shaw and Simon (1963) describe programs that could play chess and solve logic and other problems—tasks previously considered to be only in the human domain. Other famous examples from that era include demonstrations that the output of computers could be confused with human performance. For example, ELIZA (Weizenbaum, 1966) was an "artificial intelligence" program that simulated therapists interacting with patients. Some Rogerian therapists were fooled into believing that the simulations were actually other therapists responding at the other end of a terminal.

The challenges for models of human intelligence became stiffer in the 1970's when researchers began adding the requirement that these programs not only produce output similar to humans' problem-solving *products* but also reproduce patterns of behavior consistent with the *processes* believed to enable human performance. For example, the general problem solver (GPS) was a computer program designed to solve problems in a manner consistent with a problem-solving heuristic used by humans (Newell & Simon, 1972). This model and others from its era were tested against human data by evaluating the degree of match between the intermediate steps taken by the model and those taken by humans tackling the same problems. Thus, the mode of comparison was generally qualitative (e.g., Do the model's steps indicate the same processes used by humans?) rather than quantitative.

Following the successes of these models, researchers in the 1980's aimed to attain even more stringent fits to the observed data by getting their models to match quantitative measures such as accuracy or latency data. For example, Seidenberg and McClelland (1989) implemented their theory of visual word recognition in the form of a connectionist model and showed that the model provided a good fit to the data from several experiments. In this case, the model's fit was assessed by correlating a quantitative measure collected in the experiments (reaction time) to a corresponding but different quantitative measure produced by the model (squared error). A similar level of model fit was achieved by Anderson's (1983) activation-based model of the fan effect. Here, the model was able to fit participants' response latencies for various kinds of memory-verification stimuli. In these two cases as in others from the 1980's, the processing speed of computers available enabled modelers to automatically search the space of free parameters and present the results from the best fitting parameterization of their model. Hence, models of this era were generally evaluated based on their ability to fit a relatively large number of data points with a fixed number of freely varying parameters; when such models were fitted across experiments, independent sets of parameters were typically used.

The move toward correlating models' quantitative predictions with corresponding data made it much easier to rigorously evaluate the fit of a given model as well as to compare the fit of different models (e.g., by correlation or goodness-of-fit statistics). However, with numerous free parameters and only univariate data to be fit, multiple models implementing very different theories could potentially produce fits of equivalent quality. Hence, the race was on to demonstrate a model's quality of fit in additional ways. In the 1990's, researchers applied extra requirements for their models to meet. For instance, models in this era were often evaluated based on the quantitative fit of *multiple* dependent measures. An example of this is given in Anderson, Bothell, Lebiere, and Matessa (1998), where a model of list memory was fit to participants' latency and accuracy data simultaneously across several conditions that varied the list size and recall order.

In other cases, models were extended so they could go beyond fitting aggregate data; these models were designed to be able to vary their performance (e.g., by continuously varying one or more parameters) in order to capture variation across a population. Many of these modeling efforts achieved good fits to data from various subgroups of participants in the population (e.g., Just & Carpenter, 1992; Byrne, 1998). More rare was the attempt to capture particular *individual* participants' behavior (e.g., Nosofsky & Palmeri, 1997; Ratcliff, Van Zandt, & McKoon, 1999). For example, Nosofsky and Palmeri (1997) collected data from three participants performing a color classification task and then showed that their model of the task was flexible enough to match each individual's pattern of performance by allowing six model parameters to vary.

Note that in each of the above cases from the 1990's, researchers were improving their models in ways suggested (and tested) by the data. Multiple measures, variability in a dataset, and subgroups' or individuals' patterns of performance are all aspects of the data that can be used to guide model development. And yet, no model is perfect; it is only an approximation of the true processes under study. So, the more computational models being advanced, the more these extra data-based constraints helped by limiting the space of candidate models to those best able to simulate the desired processes in as many ways as possible.

Yet another extension to computational modeling that was advanced in the 1990's involved developing models within *cognitive architectures* (e.g., ACT-R: Anderson & Lebiere, 1998; EPIC: Meyer & Kieras, 1997a, 1997b; Soar: Newell, 1990). A cognitive architecture is a computational system that specifies a particular way of representing information and a fixed set of cognitive mechanisms for processing that information. Modeling within a cognitive architecture provides a different source of constraint—one that does not come from the data to be fit (as in the extensions presented above) but rather one that is imposed "top down" by the theory of the given cognitive architecture. Specifically, any model built within a cognitive architecture must use the same representations and mechanisms regardless of the task being modeled, just as the brain presumably employs a common set of information-processing mechanisms across a variety of tasks. What differs across various models built within a given architecture is the task-specific knowledge with which each task model is endowed. In sum, the goal of this "architectural" modeling approach was to develop models of a variety of different tasks under a single architecture and to show that, while using the same set of information-processing mechanisms, these models all fit their respective datasets. This goal has been achieved to varying degrees by researchers working within the cognitive architectures mentioned above.

Although the above modeling accomplishments show enormous progress, there are still areas where the full potential of computational modeling has not yet been achieved. One such area involves the modeling of individual differences in cognitive processing, i.e., processing capabilities that differ among individuals but that are relatively constant within individuals as they work on a variety of tasks. Just as the human brain allows for the performance of many tasks using a set of mechanisms that are presumably shared by many individuals, it also allows for considerable variation among individuals in the quality and speed of task performance. Computational models need to be able account for both the commonalty among individuals' processing (e.g., a common set of mechanisms for learning and performance, regardless of the task or the person) as well as the variation in individuals' processing (e.g., differences in the fundamental processing capacities with which these shared mechanisms are run). For example, cognitive models have not yet been developed that predict the performance of individual participants across tasks and along multiple dimensions, even though their performance is related across those tasks. Ideally, such a modeling effort would be able to predict individuals' performance in a new task with no new free parameters, presumably after deriving an estimate of each individual's processing parameter from previous modeling of other tasks. We take this as a challenging but feasible goal that may help bring computational models to the next step in their progression.

Computational Modeling and Working Memory

Working memory is one area of cognitive processing where systematic individual differences have been found experimentally and where computational models have already made some progress toward the above goal of predicting individual performance. Working memory is the set of mechanisms used in human cognition for retrieving and maintaining information during processing (Baddeley, 1986, 1990). For example, to compute the proper amount to tip in a restaurant, working memory resources are required to hold the original bill amount and any intermediate results in memory while working toward the final answer. Because working memory resources are limited, performance suffers when the working memory demands of a task exceed the available supply. Indeed, prior research has demonstrated that as the working memory demands of a task increase, people's performance on the task decreases (e.g., Anderson & Jeffries, 1985; Anderson, Reder, & Lebiere, 1996; Baddeley, 1986; Burgess & Hitch, 1992;

Engle, 1994; Just & Carpenter, 1992; Navon & Gopher, 1979; Salthouse, 1992). Salthouse, for instance, had participants perform four different tasks at three levels of complexity. He found that as task complexity increased, performance decreased. Salthouse also found individual differences in performance such that the decrease in performance with increased task complexity was greater for older adults. Other researchers have also found that people differ in their sensitivity to the working memory demands of a task, regardless of their age (e.g., Cantor & Engle, 1993; Carpenter, Just, & Shell, 1990; Conway & Engle, 1994; Engle, 1994; Just & Carpenter, 1992). Moreover, these individual differences in working memory capacity appear to explain a substantial amount of the commonalty in an individual's performance across tasks. Engle (1994), for example, reported that an individual's working memory capacity (as measured by performance on a specially designed task) correlates well with performance on a variety of other tasks including reading, following directions, learning vocabulary and spelling, notetaking, and writing. Engle interprets this correlation as evidence that all of these tasks require use of a common resource, the individual's working memory, which influences performance. In sum, working memory capacity, it seems, is a resource that limits performance on highly demanding tasks, differs in amount across individuals, and helps predict individuals' performance across different tasks. Any complete model of working memory must produce all three of these phenomena.

Several approaches have been taken to model working memory. Some theorists have used statistical methods (e.g., factor analysis) to look for clusters of abilities that tap working memory (e.g., Carrol, 1993; Kyllonen, 1995) while others have offered qualitative descriptions of proposed working memory mechanisms (e.g., Baddeley, 1986; Salthouse, 1996). Still others have offered computational models of working memory performance. For example, Burgess and Hitch (1992) developed a connectionist model of Baddeley's (1986, 1990) articulatory loop, a component of working memory hypothesized to hold verbal stimuli for a limited amount of time. In their model, item-item and item-context associations are learned via connection weights which propagate activation between memory items and enable sequential rehearsal through a list. Because these weights decay with time, more demanding tasks (e.g., those requiring the maintenance/rehearsal of longer lists or lists of longer words) tend to propagate less activation to the memory items, leading to more errors. Another connectionist model of working memory takes a different approach: O'Reilly, Braver, & Cohen (1999) have proposed a biologically inspired model in which working memory functions are distributed across several brain systems. In particular, their model relies on the interaction between a prefrontal cortex system, which maintains information about the current context by recurrently activating the relevant items, and a hippocampal system, which rapidly learns arbitrary associations (e.g., to represent stimulus ensembles). Both systems' excitatory activation processes, however, are countered by inhibition and interference processes such that only a limited number of items can be simultaneously maintained. This limitation leads to decreased performance in complex tasks. Both of these models have been fit to aggregate data. That is, they have captured one of the main working memory results, namely, that performance suffers as the working memory demands of a task increase.

Another approach to working memory is taken by EPIC, a production-system architecture (Kieras, Meyer, Mueller, & seymour, 1999; Meyer & Kieras, 1997a, 1997b). Here, an articulatory loop is implemented via the combined performance of an auditory store, a vocal motor processor, a production-rule interpreter, and various other information stores. Performance of the model is governed by production rules which implement strategies for rehearsal and recall and which, in turn, draw on the capabilities of the other components. Working memory limitations stem from the all-or-none decay of items from the auditory store (with time until decay being a stochastic function of the similarity among items) and from the articulation rate attributed to the vocal motor processor. As the vocal motor processor takes the prescribed amount of time to rehearse a given item (re-adding that item to the auditory store), other items have a chance to decay (disappearing from the auditory store), thereby producing subsequent recall errors. This model also accounts for the performance decrease with increased working memory demands, and it does this within a cognitive architecture.

These models have accounted for a wide range of working memory effects at the group level (i.e., aggregated across participants). As noted previously, however, individuals differ in their working memory capacity. These capacity differences can result in very different patterns of performance by individual. A complete theory of working memory should, therefore, be able to capture not only aggregate working memory effects but also the differential sensitivities of individual participants to working memory demands, i.e., the second working memory result mentioned above. Just and Carpenter (1992) developed a model called 3CAPS that was able to capture the different patterns of performance in subgroups of participants with low, medium, or high working memory capacity. Their model accounted for these differences by assuming different caps on the total activation propagating through the system, Thus, the 3CAPS model accounted for individual differences in working memory capacity at the sub-group level.

An ACT-R Model of Working Memory

In this section, we trace the development of a computational model of working memory built within the ACT-R architecture. As the current implementation of ACT-R stands (cf. Anderson & Lebiere, 1998), it already offers an account for the working memory result that participants' aggregate performance degrades as tasks become more demanding. To account for the other two critical working memory phenomena, which both deal with individual differences in working memory, we extend the theory. Specifically, we posit that the continuously valued parameter W represents an individual's working memory capacity. In our extended model, then, the value of W can be varied to produce the different patterns of performance that are observed for different individuals, i.e., the second working memory result mentioned above. More importantly, however, we show that by using the model parameter W as a measure of an individual's working memory capacity, we can predict that individual's performance on a new task without otherwise tailoring the model. This is an example of the third working memory result from above that states there is a common working memory resource influencing an individual's performance across tasks. Thus, our model satisfies all three requirements of a good working memory model while, at the same time, fitting multiple dependent measures at a fine-grained level of detail.

ACT-R Fundamentals

ACT-R (Anderson & Lebiere, 1998) is a cognitive architecture and, as such, it provides the basic mechanisms for how computational models built within it control cognitive processing and store and retrieve information. ACT-R models of individual tasks specify the declarative and procedural knowledge required to perform a specific task, but that knowledge is always processed, stored, and retrieved according to the fixed set of ACT-R mechanisms. Moreover, declarative and procedural knowledge are each represented in a particular way, determined by the ACT-R architecture. Declarative knowledge is represented as schema-like structures called chunks. A single chunk consists of a node of a particular type and some number of slots that encode the chunk's contents. Figure 1 represents the memory for the fact that seven is the item in

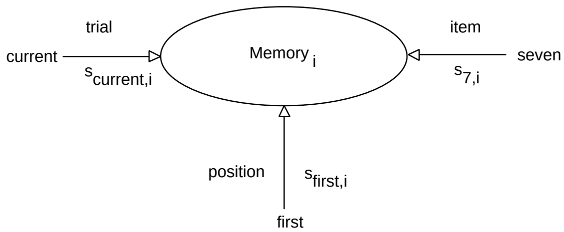

Figure 1. Graphic representation of a chunk encoding the fact that seven was the first item in the current list.

the first position of the current trial. Retrieval of such a chunk is based on its total activation level. This level is determined by the chunk's base-level activation (derived from its history of use) and the amount of source activation it receives from elements currently in the focus of attention (representing contextual effects). Chunk activation is given by:

$$A_i = B_i + \sum_{j=1}^{n} \frac{W}{n} S_{ji} \qquad \text{Equation 1}$$

where A_i is the total activation of chunk I, B_i is the base level activation of chunk I, W is the amount of available source activation (to be discussed in more detail below), and n is the number of elements in the focus of attention. Each S_{ji} is the strength of association between chunk j in the focus of attention and chunk i in declarative memory. In Figure 1, $S_{7,i}$ represents the link between the concept 7 and the memory of seven in the first position of the current trial. As this link becomes stronger, more source activation will be spread to the memory chunk should seven become the focus of attention. This in turn makes it easier to access and use that memory chunk in processing. In general, a chunk will be more active the more often it is used (increasing B_i) and the more strongly related it is to items in the focus of attention (higher S_{ji}'s).

The exact value of a chunk's total activation (A_I) is important because it determines the probability of the chunk being retrieved, as given by the following:

$$\text{Probability of retrieving chunk i} = \frac{1}{1 + e^{-(A_i - \tau)/s}} \qquad \text{Equation 2}$$

where A_i is, as before, the total activation of chunk i, τ is the retrieval threshold, and s is a measure of the noise level affecting chunk activations.

If a chunk's total activation (plus added noise) is above the threshold τ, the chunk is retrieved and its latency of retrieval is given by the following:

$$\text{Latency to retrieve chunk i} = Fe^{-fA} \qquad \text{Equation 3}$$

where F and f are constants mapping A_i onto latency. If the total activation falls below threshold, the model commits an error of omission. Errors of commission are produced by a partial matching mechanism that is described in more detail in Anderson and Lebiere (1998).

In recent work, Anderson, Reder, and Lebiere (1996) suggested a computational approach to working memory within ACT-R that builds on the work of Just and Carpenter (1992). While Just and Carpenter proposed that the *total* activation within the cognitive system is limited, Anderson et al. (1996) suggested that working memory limits occur because source activation (the parameter W in Equation 1) is limited (see Cantor & Engle, 1993 for a similar account). Source activation is a type of attentional activation that is divided equally among the items in the current focus of attention. It spreads from these items to related chunks that are necessary for task performance and in this way maintains those task-relevant chunks in an available state relative to the rest of declarative memory. Because the amount of source activation is limited to the quantity W and because this quantity is divided among the various items in the current focus of attention, the more items in the focus, the less source activation each can spread to its related chunks. For example, increasing the complexity of a task (which increases the number of items in the focus of attention) implies that each item in the focus of attention has a smaller share of source activation to spread to task-relevant chunks. Similarly, when a person is dividing attention between two concurrent tasks (which places more items in the focus of attention than would be for either task alone) there will be less source activation spreading from each item in the focus to task-relevant chunks. In both cases, the task-relevant chunks are less activated than they would be in a simple, single task, and so performance suffers as a result.

Anderson et al. (1996) supported this conceptualization of working memory by demonstrating that an ACT-R model with a limit on source activation provided a good fit to their participants' data. In their study, participants were required to hold a digit span in memory while solving an equation. Both the size of the span and the complexity of the equations to be solved were manipulated. Also, half of the equations included extra variables such that participants had to substitute the first two values from the digit span for these variables before solving. Results showed that when the equation-solving task was made more complex (i.e., more operations were required to solve), performance on both tasks decreased. When the span task was made more complex (i.e., more digits to hold in memory), performance on both tasks also decreased whenever the equation required substitution from the digit span.

Extending the ACT-R Theory

Like Anderson et al. (1996), we assume a fixed limit on W from Equation 1. This assumption accounts for the first of the three characteristics of working memory outlined above, namely, that working memory resources, represented by W, are inherently limited. Note that a limit on W will affect the performance of any ACT-R model where task performance relies on retrieval of declarative chunks. This is because declarative chunks are influenced by the amount of source activation (W) spreading from the current focus of attention (Equation 1).

To account for the second working memory phenomenon— that these resources differ in amount across individuals— we extend the idea of a fixed "cap" on source activation, W. Specifically, we assume that the limit on W is not the same for each individual, i.e., that different participants will have different W values. Moreover, we expect that the distribution of these W values across a population follows a normal distribution centered at 1.0. In this way, different versions of an ACT-R model can be endowed with different amounts of source activation to represent different individuals. The higher the value of W, the higher task-relevant chunks' total activations will be relative to the rest of declarative memory (Equation 1). These activations in turn will impact the probability and speed of successful retrieval of the task-relevant chunks (Equations 2 and 3). Hence, our working memory model predicts that an individual with a larger value of W will be able to retrieve task-relevant information more accurately and more quickly than will an individual with a smaller value of W. Moreover, a high value of W (making W/n large) will only lead to degraded performance (let activation A fall below threshold τ) when the task is sufficiently complex (n is sufficiently large), whereas a low value of W will lead to degraded performance at much lower levels of complexity.[1] This allows the model to produce differential sensitivity to the working memory demands of a task.

Finally, to account for the third working memory characteristic, namely that an individual's performance profiles across different tasks will be related because of his or her particular level of working memory resources, we take total source activation (W) to be relatively constant within each individual—even across different tasks. Because source activation spreads the same way in any ACT-R model, our view of working memory can produce

[1] This is assuming all else equal (e.g., τ, s, S_{ij}, B_i, etc.)

the above effects (e.g., differential sensitivity to working memory demands) across tasks. It is, of course, possible that an individual's W will vary with variations in the individual's level of interest, degree of practice, or with fatigue, but we take these variations to be smaller than the variability among individuals.

Testing the Model

In this and the next section, we document how we have tested the above model of working memory in the context of two particular working memory-dependent tasks. Before doing so, however, we carefully selected those tasks and then specified the knowledge required for performing each. This task-specific knowledge was represented in terms of ACT-R's declarative chunks and production rules. Then, ACT-R's general mechanisms—including our extension regarding how to represent individual differences in working memory—was applied to each task model in order to generate specific predictions for how participants would perform that task. We then compared these predictions with the data from our participants, both at the aggregate and individual level.

Challenges to Modeling Working Memory

Selecting a task that will enable estimation of an individual's working memory capacity is challenging because performance on a given task can be affected by a number of other factors, including prior knowledge of relevant procedures and possession of related facts. A suitable working memory task should limit the inter-participant variability on these other factors. In traditional memory span tasks, participants are presented a sequence of stimuli (i.e., digits, letters, words) one at a time and then are required to repeat the sequence. Successive sequences are lengthened until the participant can no longer repeat them accurately. Working memory capacity is taken as the longest sequence that a participant can accurately report. Such tasks, however, do not allow for very accurate measures of working memory capacity because the use of compensatory strategies (i.e., chunking, mnemonics) has been shown to differ among participants, and such differences in strategies can seriously contaminate measures of working memory capacity derived from such tasks (Turner & Engle, 1989). A further concern is the influence of task-relevant factual knowledge on performance of such tasks. To cite an extreme case, Chase and Ericsson (1982) described a participant with a digit span of more than 80 digits (compared to a normal span of approximately 7 items). This feat was accomplished in part because the participant, a runner, was able after extensive practice to recode the digits into groups based on personally meaningful running times. Thus, his super-high memory span has been mainly attributed to variation in knowledge and experience rather than to variation in working memory resources alone. In sum, to obtain valid measures of working memory capacity it is critical that we find tasks that can be completed in only one way and are equally unfamiliar to all the participants.

One way to deal with the challenge of minimizing differences in compensatory strategies and prior knowledge is the use of modified span tasks in which participants perform some other activity concurrently with the test of memory span (e.g., Daneman & Carpenter, 1980; Turner & Engle, 1989; Yuill, Oakhill, & Parkin, 1989). Modified span tasks put a greater continuous load on working memory and tend to preclude participants from inventing and using different strategies that could obscure differences in working memory capacity. Lovett, Reder, and Lebiere (1999) developed and refined a *mo*dified *d*igit *s*pan (MODS) task that is a variant of one developed by Oakhill and her colleagues (e.g., Yuill, Oakhill & Parkin, 1989).

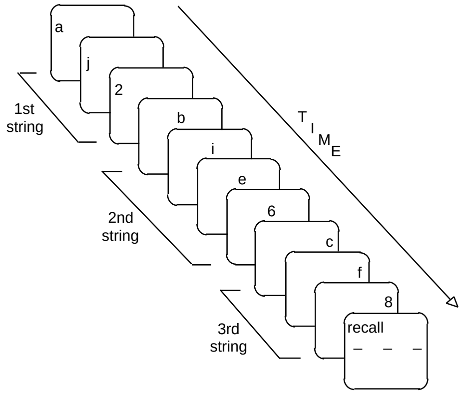

Figure 2. Graphic depiction of the MODS task. Subjects read the characters aloud and attempt to retain the digits for later recall.

In the MODS task, each trial consists of a sequence of strings, presented one character at a time on the computer screen. Each string contains a variable number of letters and ends in a digit (see Figure 2). Participants read each character aloud as it appears, and they must maintain the string-final digits in memory. Trials vary systematically in the number of strings (i.e., digits to be recalled), called the set size, from three to six. At the end of each trial, a recall prompt appears, indicating that the participants should enter the digits from that trial *in the exact order they appeared*. Participants' performance then is evaluated according to their accuracy in recall.

The MODS task fulfills both of our criteria for a suitable working memory task. First, the characters in each string are presented at a rapid pace (two characters per second), so the requirement to read these characters aloud as they appear essentially prevents participants from engaging in various strategies that might differ among participants. The requirement that participants respond verbally also acts as a type of articulatory suppression, which prevents rehearsal within each string (Baddeley, 1986). Second, we can assume that all of our participants are equally familiar with letters and numbers and that the rapid pace of the task prevents recoding that would make use of idiosyncratic chunking (e.g., recoding the digits as local telephone exchanges). As a result, this task is expected to yield relatively pure estimates of working memory capacity.

Empirical Support for Our Model—Task #1

Lovett, Reder, and Lebiere (1999) developed an ACT-R task model of the MODS task that contains the facts and procedures required for performance. For example, Figure 1 depicts our representation for a single memory element. The required procedural knowledge includes the various skills involved in this task, e.g., reading letters and digits aloud, storing the current digit as a chunk, and recalling stored chunks in order according to their positions (See Lovett et al., 1999, for more details). This model was then run under ACT-R's general mechanisms, including the new working memory model described earlier.

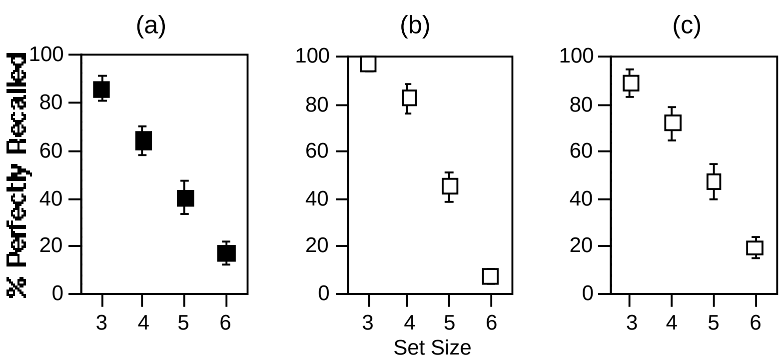

Figure 3. Data (a) and two model fits from Lovett et al. (1999), one without any variation in the simulations' working memory resources (b) and another with this variation (c).

Figure 3a shows, for each set size, the average proportion of trials recalled correctly by the 20 participants who performed the task.[2] To generate an initial set of model predictions, Lovett et al. ran 20 simulations to correspond to these 20 participants, without optimizing the

[2] The Lovett et al. study used a finer breakdown of conditions by crossing two other factors with set size. Although our model captured the variation across the additional conditions, we present the collapsed fits here for simplicity.

parameters to fit the data. In particular, each simulation had W set at its default value of 1.0. Figure 3b shows that the model's performance, averaged over these 20 simulations, follows a decreasing trend similar to the participants' aggregate performance data. However, these 20 simulations all had the same W value. To account for the variability in working memory capacity across participants in the sample, Lovett et al. ran an additional 20 simulations, but this time the value of W for each run was allowed to vary around the value 1.0. These separate W values for each simulation were not optimized to the data in any way; they were merely allowed to vary randomly from run to run. Figure 3c shows that, with these varying W's, the model's average predictions produced an even better fit to the aggregate data. This change in the absolute level of the predictions occurred because of the nonlinear effects of W on performance: Varying W across simulation runs changed not only the variability in the model's performance (notice the standard error bars in Figure 3c compared to those in 3b) but the absolute level of performance as well. Finally, Lovett, et al. were able to fit the data of *individual* participants by allowing W to take on particular values chosen to maximize the fit to individual participants' performance. This last result suggests that our model – even when constrained to let only one parameter vary among participants – is flexible enough to capture individual differences in performance.

In two recent experiments involving the MODS task, Daily, Lovett, and Reder (1999) have used this same modeling approach to capture individual differences in working memory. Replicating the results of Lovett et al. (1999), Daily et al. found that running the model with W varying about the default value of 1.0 produced a good fit to the aggregate data (Figure 4). Moreover, by leaving all parameters fixed except for estimating a best-fitting W value for each participant, the model was able to capture the particular pattern of performance exhibited by each individual. Figure 5 shows individual participant fits for four participants chosen to represent a range of W values.

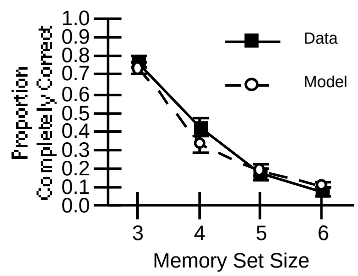

Figure 4. Model fit to the aggregate data in Daily et al (1999).

Going beyond the Lovett et al. results, Daily et al. tested their model's ability to predict each participant's serial position curves. Without changing any additional parameters of the model, they generated, for each participant, the model's predictions of the proportion of trials in which a given memory position in a given set size would be recalled correctly. For example, on what proportion of trials of set size three were the first, second, and third digits recalled correctly? Figure 6 shows these model fits for set size six for the same four participants as in Figure 5. (Set size 6 is only included here to avoid too many overlapping curves in each panel.) It is clear from these plots that the model can capture detailed differences in participant performance of the MODS task using W as the only free parameter.

These results demonstrate that our model is able to capture human performance in a working memory task at three levels: set size effects averaged across participants, set size effects within individual participants, and serial position effects within individual participants. Further, to the best of our knowledge, this is the first demonstration of a computational model predicting individual participant performance using a single parameter to capture individual differences.

Cross-task Prediction of Performance

Above, we showed that our working memory model's single individual difference parameter can capture individual participants' patterns of performance in the MODS task. In this section, we demonstrate that, in addition, estimates of W derived from individual participants' performance on the MODS task lead to accurate predictions of their individual performance on a second working memory task (with all other parameters held constant). This adds to the support for W as a measure of working memory capacity and offers the first demonstration of modeling individual differences with a single free parameter within the first task and no free parameters in the second task.

Issues in Choosing Task #2

The second task we chose for this cross-task prediction of individuals' performance is the n-back task (e.g., Braver et al., 1997; Cohen et al, 1994; Gevins, et al., 1996; Gevins, Smith, McEvoy, & Yu, 1997). In this task, participants are presented with a long sequence of letters and are required to indicate whether the current letter matches some previous letter in the sequence. For instance, in a 1-back condition, the participant is told to respond positively when the current letter matches the immediately preceding letter in the sequence. In a 2-back condition, the participant is told to respond positively when the current letter matches the second letter before the current one, etc. Thus, as the number of items "back" increases, the participant must keep track of a greater number of items in order to respond accurately. In addition, a 0-back condition is often included in which participants must indicate whether the current letter matches a fixed letter. The usual finding in the n-back task is that response latency increases and accuracy decreases as memory load (i.e., the value of "n") increases (Braver et al., 1997; Cohen et al, 1994).

For our purposes, we collected data from 20 participants performing both the MODS task (as described earlier) and four conditions of the n-back task (0-back, 1-back, 2-back, and 3-back). Our modeling approach then involved fitting individual participants' data from the MODS task to estimate individual W values and then using these W values to generate individualized predictions for these same participants' nback data.

Note that the n-back task is qualitatively different from the MODS task in several respects. Unlike the MODS task, which requires *recall* of the memory set items, the n-back task involves *recognition* of previously presented items. Further, successful performance of the n-back task requires continual updating of a stream of to-be-remembered items whereas the MODS task involves maintenance of a separate list of to-be-remembered items for each trial. Finally, the

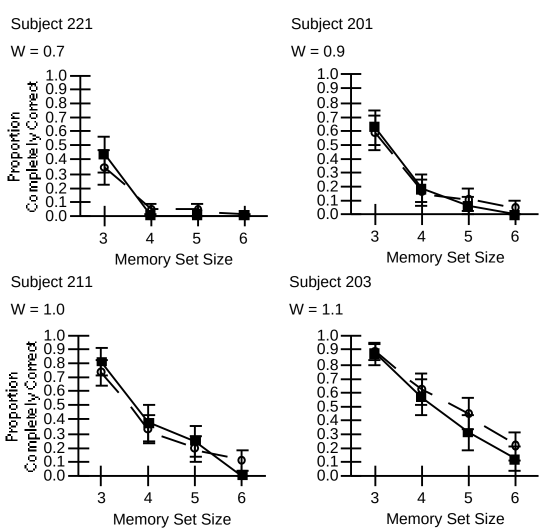

Figure 5. Model fits for four representative subjects from Daily et al. (1999). Filled symbols are subject data, open symbols are the model's predictions.

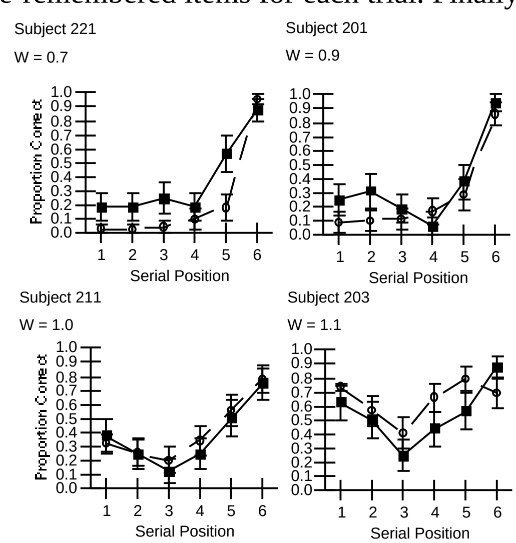

Figure 6. Fits to the serial position data for 4 typical subjects (largest set size only). Filled symbols are subject data, open symbols are the model's predictions.

memory load involved in the n-back (three items at most) is somewhat smaller than the load in the MODS task (varies from 3 to 6 items). Given these differences, our working memory model's ability to make cross-task predictions based on a single individual difference parameter suggests that there is something general about the impact of source activation (W) on performance.

Perhaps more important than the n-back task's differences from the MODS task is its status regarding our two criteria for a suitable working memory task. Our original assumption was that participants would not differ in terms of prior knowledge relevant to this task nor in the strategy they used for task performance. To test these assumptions, we included a questionnaire asking participants how they performed the n-back task. Participants' responses suggested that they were equally unfamiliar with the task, i.e., participants did not differ in relevant prior knowledge. Contrary to our expectations, however, participants indicated using one of two qualitatively different strategies. In the first (which we named the *activation* strategy), participants responded to each letter based on its familiarity: if the item seemed familiar it was called a match. The second strategy (which we named the *update* strategy) involved actively maintaining a list of the prior letters and updating that list after each letter was presented. Presumably, working memory resources would not be involved in the first strategy as no maintenance is involved. These resources would, however, be required for the maintenance and updating of the lists in the second strategy. Therefore, we chose to model the nback data from the update group only by developing an ACT-R model to implement the update strategy.

Empirical Results for Task #2

We divided the participants into two groups based on their self-reported strategy and compared the performance of the two groups (see Siegler, 1987). Participants' proportion correct as a function of memory load is shown in Figure 7. These data were entered into a 2 (Group) by 4 (Memory Load) analysis of variance. The effect of group was marginally significant, $F(1, 18) = 2.86$, $p = .108$, MSE = 0.01. The update group tended to be more accurate than the activation group. The effect of memory load was significant, $F(3, 54) = 142.98$, $p < .001$, MSE = 0.18. Participants were less accurate as load increased. Finally, the interaction between group and memory load was also significant, $F(3, 54) = 2.84$, $p = .046$, MSE = 0.003. While accuracy decreased for both groups as load increased, the decrease was smaller for the update group than for the activation group. This difference in performance suggests that the differences in the strategies adopted by the two groups are not trivial and that the strategies have a real affect on performance.

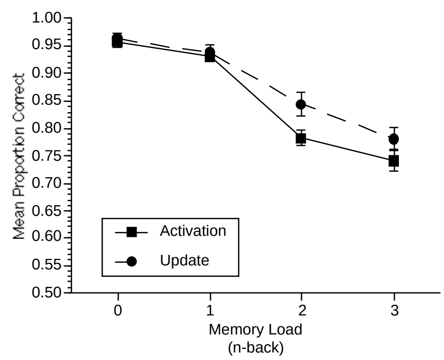

Figure 7. Accuracy in the n-back task as a function of the strategy adopted by subjects.

As mentioned earlier, we expected that working memory resources would not be involved in the activation strategy but would be in the update strategy. To garner some preliminary evidence for this distinction before modeling the data, we divided the participants into two groups according to their stated strategy and, separately for each group, tested the correlation between W, as estimated from participants' MODS performance (see below) and their overall n-back accuracy. Consistent with our expectations, we found that that there was no relation between W and n-back accuracy in the activation group ($r = 0.04$, $p = .92$) and a marginally significant relation between W and accuracy in the update group ($r = .56$, $p = .12$).

Although this second result is not highly significant, it is still consistent with our expectations. This is because our working memory model does not predict a stictly linear relationship between W and performance; rather, the relationship is nonlinear and likely involves the complexity of both direct and indirect effects. Thus, some degree of correlation was sufficient encouragement to proceed with the modeling of the update group.

One concern, however, was that this selection of the update group might have restricted or shifted the range of working memory capacities to be modeled. To explore this we compared the two groups of participants in terms of their estimated W values and found no difference: Both groups had a mean W of 0.97 with standard deviation 0.09, suggesting that participants' choice of strategy does not relate to working memory capacity.

Modeling Results for Task #1

Individual W estimates were computed from each participant's MODS data using the same model and the same parameters as in Daily et al. and merely estimated a new W value for each participant in our new sample. Individual MODS data and model fits from four representative participants, chosen to represent a range of W values, are shown in Figure 8. As in Daily et al. (1999), these fits are quite good. In addition, we used our MODS model with the same set of individually estimated W values to generate predictions of individual participants' serial position curves. Figure 9 shows the data and model predictions for the set size 6 serial position curves for the same participants as in Figure 8. Given the quality of these individual participant fits, we moved on to modeling participants' nback data.

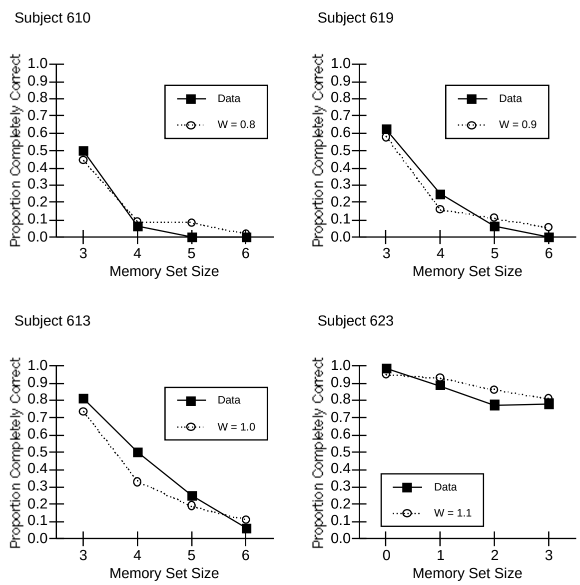

Figure 8. Fits of the MODS model to data from 4 representative subjects. Global ACT-R parameters for these fits were set in Daily et al. (1999).

Aggregate Modeling Results for Task #2

As mentioned above, we developed an ACT-R task model for the nback task that included all the declarative and procedural knowledge necessary for performing this task according to the update strategy. In this model, each stored letter was represented as a declarative chunk indexed according to how many letters back it was from the current letter. The procedural knowledge included a set of rules for reading the letters on the screen, updating the positions of the letters currently being maintained, and testing whether the current letter matched the memory for the nth-back letter.

Figure 10 shows that this ACT-R task model was able to fit the aggregate nback data ("update" participants only). For this fit, most of ACT-R's global parameters were left at their default values (see Anderson & Lebiere, 1998). Activation noise (the s in Equation 2), which has no default value in ACT-R, was pre-set at the arbitrary value of 0.04 (the same value used in the MODS model). Retrieval threshold (τ in Equation 2) was the only parameter estimated to optimize the fit to the observed data. Its optimal value was 1.80. The fit shown in Figure 10 is quite good with $R^2 = .99$.

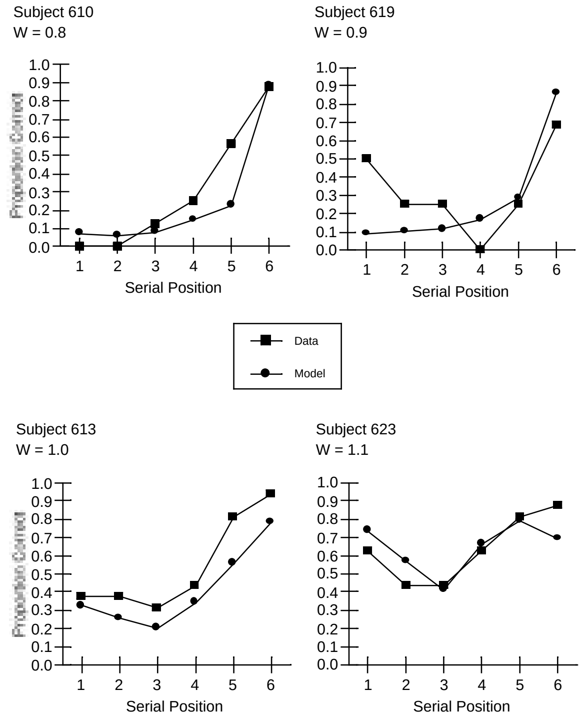

Figure 9. Individual serial position fits (largest set size only) for the same subjects shown in Figure 8.

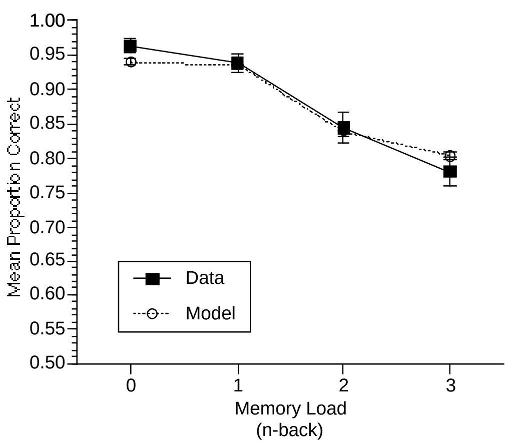

Figure 10. Aggregate data and model fit from the n-back task.

Predicting Individuals' Performance in Task #2 from Modeling Task #1

The real question in fitting these nback data was whether the W values estimated for each participant from the MODS task would predict those participants' *individual* performance data on the n-back task. To answer this question, the W value for each participant (estimated from fitting to the MODS task model) was simply plugged into the n-back model to obtain predictions of that participant's n-back performance. That is, no new parameters were estimated to fit these individual participant data. The model's predictions for four representative participants are presented in Figure 11. As with the individual participant fits for the MODS task, these fits are quite good. Over all participants' individual nback data points, the model fit has R^2=.79. This quality of fit is as good as that obtained from predictions produced by optimizing an entire new set of W values to fit the individual nback data themselves. This suggests that the W values estimated from the MODS data are capturing the important systematic variation in the nback data, i.e., that W captures a common working memory resource. In contrast, if we fit each participant's nback data with ACT-R's default W of 1.0, the corresponding model fit has a lower R^2 value of .68.

We want to emphasize that these fits were obtained in a three step process with only one free parameter per participant in the first step, one additional global free parameter in the second step, and zero free parameters in the third step. For the first step, we used our previously tested model of the MODS task to fit the MODS data of our new participants. The model and its global parameter settings were used unchanged; we only estimated one new W value for each new participant. For the second step, we fit our new nback model to the aggregate nback data. Here, we fixed W at 1.0 and only estimated one free global parameter, the retrieval threshold. For the third step, we fixed the retrieval threshold at this optimal value and used the nback model with the MODS-based estimates of each participant's W value to generate predictions for each participant's nback data. Specifically, we fit these individual participant data using only parameters that were either set at their default values or estimated in a previous step, i.e., zero

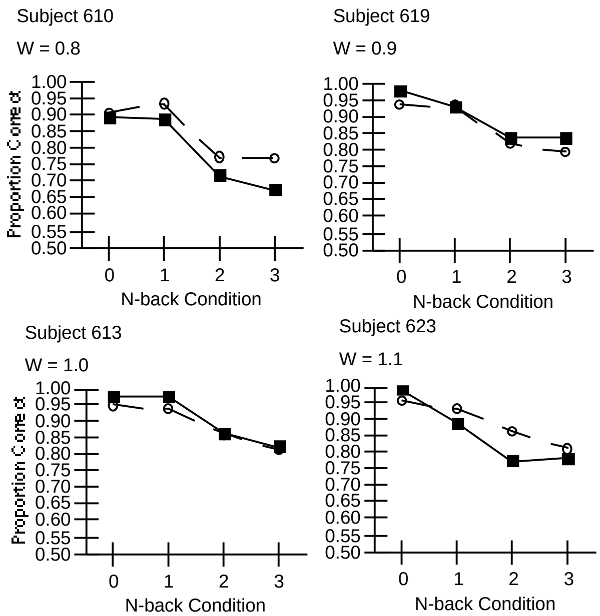

Figure 11. Individual subject fits from the n-back task. W values were estimated from subjects' MODS performance.

additional parameters.This procedure constitutes a type of generalization test (see Busemeyer & Wang, in press): parameters estimated from one task are used to make a priori predictions concerning new experimental conditions, providing a test of the model's ability to accurately extrapolate. In sum, focusing on the nine participants for whom we modeled both MODS and nback performance data, we estimated a total of 10 parameters to fit a total of 72 data points across two tasks.

Discussion

We have presented a theory that conceives of working memory capacity as a limit on source activation, a specific kind of activation used to maintain goal-relevant information in an available state. As the demands of a task increase (i.e., the number of items that must be maintained increases), this limit results in degraded performance. The key element to our theory is the hypothesis that the amount of source activation varies among individuals and that these differences account for differences in performance of working memory tasks, provided that prior knowledge of task-relevant facts and procedures are controlled. We implemented this theory in a computational model built within the ACT-R cognitive architecture (Anderson & Lebiere, 1998) and demonstrated that we could capture different patterns of performance exhibited by individual participants by varying only ACT-R's source activation parameter, W. Further, we showed that source activation could be estimated from one task and used to produce accurate zero-parameter fits to individual participant data from a second, qualitatively different task. The findings presented here strongly support this new, mechanistic view of working memory and provide evidence that our general approach is feasible.

Comparison to Other Working Memory Models

Our conceptualization of working memory as source activation is closely related to other resource-based views of working memory capacity (e.g., Engle, Kane, & Tuholski, in press; Just & Carpenter, 1992; Shah & Miyake, 1996). There are differences among the various resource-based views, however. Our theory, for example, postulates a limit on source activation, a specific kind of dynamically changing activation linked to the current focus of attention. In contrast, Just and Carpenter model working memory capacity as a limit on the total activation in the system. Another difference among these resource views involves the degree to which working memory is considered a general resource (e.g., Engle, 1994; Engle et al., in press) or a set of separate modality-based or representation-based resources (as in a separate resource for maintaining spatial versus verbal materials, Baddeley & Logie, 1999; Shah & Miyake, 1996). Our model of working memory takes source activation as a general resource whose limits should affect performance on all sorts of tasks. There are, however, patterns of performance within individuals that suggest a verbal-spatial distinction. Working memory models that postulate separate modality-based resources can account for these results directly. Instead, we offer a separate, experience-based explanation based on the idea that different people likely differ in their relative use of spatial versus verbal information. According to ACT-R, these differences would be reflected as different base-level activations (B_i's in Equation 1). Different base-level activations, like different values of W, affect performance (via Equations 2 and 3), but the base-level activation differences can vary systematically across information type whereas an individual's W value would have a general effect. Moreover, the effects of W vary with base-level activations; when base-level activation is high, the modulating effect of W is small. Conversely, when base-level activations are low, the effect of W is large. Thus, our view of working memory as a single resource still admits a variety of modality- and representation-based differences in within-individual performance.

Other Sources of Individual Differences

It should be emphasized that our arguments regarding individual differences do not claim that source activation, W, is the only thing that explains differences in people's performance on laboratory tasks or on everyday tasks. As noted above, individual differences in prior knowledge and strategies can significantly influence task performance (Chase & Ericsson, 1982; Ericsson & Kintsch, 1995). However, we have argued that when those other sources of variability in task performance are reduced, differences in source activation (our model's representation of working memory capacity) can largely explain differences in an individual's task performance. We have taken two approaches to reducing the influence of prior knowledge and strategies. The first approach, used in the MODS task, involves designing the task in such a way as to use knowledge presumed to be constant across participants and to eliminate the use of all but the most rudimentary strategy. The second approach, adopted for the n-back task, used participants' self-reports to determine which strategy a participant used and to include in our modeling only those participants who used a common working memory dependent strategy. The fact that W was able to accurately capture different patterns of performance in both of these tasks provides converging evidence that we were able to reduce other sources of variation, thereby highlighting the effects of individual differences in working memory capacity.

Conclusions

In describing working memory phenomena, we identified three important characteristics that any computational model of working memory must be able to produce: (i) working memory resources limit performance on highly demanding tasks, (ii) working memory resources differ in amount across individuals, and (iii) these differences help predict individuals' performance across different tasks. Our source activation theory possesses all of these characteristics and accurately captures working memory effects at both the aggregate and individual levels. We believe, therefore, that it provides a workable account of individual differences in working memory capacity. Though previous research has highlighted that individual differences exist, these differences have not been modeled at the level of the individual participant nor has the performance of a participant on one task been used to provide fine-grained predictions of that participant's performance on another task. That we were able to do so speaks to the power of our approach and to the generality of the ACT-R architecture.

In terms of computational modeling more generally, our approach parallels that of other, current efforts by achieving the following: (1) fitting aggregate data with multiple measures (generally with global parameter values taken from previous work), (2) accounting for the variability of the aggregate data (by varying only one model parameter to account for the range in participant performance), and (3) matching the performance patterns of individual participants (by estimating a single W value per participant we account for the fact that performance drops off under increased load more quickly for low W than for high W participants). Moreover, we have accomplished this with the additional constraints of (4) building our models within a cognitive architecture and (5) varying relatively few parameters to achieve our model fits. Most importantly, however, we have also demonstrated the ability of our model to (6) predict individual participants' performance across tasks using a single parameter estimated from one task to predict performance on the second. This is the first such demonstration of these capabilities and strongly suggests that computational models may be fruitfully employed in the investigation of individual differences.

References

Anderson, J. R. (1983). *The Architecture of Cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). List memory. In J. R. Anderson & C. Lebiere (Eds). *Atomic Components of Thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *Atomic Components of Thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., & Jeffries, R. (1985). Novice LISP errors: Undetected losses of information from working memory. *Human Computer Interaction, 1,* 107-131.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30,* 221-256.

Baddeley, A. D. (1986). *Working memory.* New York: Oxford University Press.

Baddeley, A. D. (1990). *Human memory: Theory and practice.* Boston: Allyn and Bacon.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* New York: Cambridge University Press.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvment in human working memory. *Neuroimage, 5,* 49-62.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31,* 429-460.

Byrne, M. D. (1998). Taking a computational approach to aging: The SPAN theory of working memory. *Psychology & Aging, 13*, 309-322.

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1101-1114.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the Raven Progressive Matrices Test. *Psychological Review, 97,* 404-431.

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

Chase, W. G., & Ericsson, K. A. (1982). Skill and Working Memory. In G. H.. Bower (Ed) *The Psychology of Learning and Motivation, 16,* p. 1-58, NY: Academic Press.

Cohen, J. D., Forman, S. D., Braver, T. S., Casey, B. J., Servan-Schreiber, D., & Noll, D. C. (1994). Activation of prefrontal cortex in a non-spatial working memory task with functional MRI. *Human Brain Mapping, 1,* 293-304.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354-373.

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

Daily, L. Z., Lovett, M. C., & Reder. L. M. (1999). A computational model of individual differences in working memory capacity. To appear in *Cognitive Science.*

Engle, R. W. (1994). Individual differences in memory and implications for learning. In R. Sternberg (Ed.), *Encyclopedia of Intelligence.* New York: Macmillan.

Ericsson, K. A.., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211-245.

Feigenbaum, E. A., & Feldman, J. (1963). *Computers and Thought.* New York: McGraw Hill.
Gevins, A., Smith, M. E., Le, J., Leong, H., Bennet, J., Martin, N., McEvoy, L., & Whitfield, S. (1996). High resolution evoked potential imaging of the cortical dynamics of human working memory. *Electroencephalography and Neurophysiology, 98,* 327-348.
Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex, 7,* 374-385.
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.
Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* New York: Cambridge University Press.
Kyllonen, P. C. (1995). CAM: A theoretical framework for cognitive abilities measurement. In D. Detterman (Ed.), *Current topics in human intelligence: Volume IV, Theories of intelligence.* Norwood, NJ: Ablex.
Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* New York: Cambridge University Press.
Meyer, D. E., Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: I. Basic mechanisms. *Psychological Review, 104,* 3-65.
Meyer, D. E., Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104,* 749-791.
Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86,* 214-255.
Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard.
Newell, A., Shaw, J. C., & Simon, H. A. (1963). Empirical explorations with the logic theory machine: A case study in heuristics. In E. A. Feigenbaum & J. Feldman (Eds). *Computers and Thought.* New York: McGraw Hill.
Newell, A., & Simon, H. A. (1972). *Human Problem Solving.* Cambridge, MA: Harvard.
Nosofsky, R. M., Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266-300.
O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* New York: Cambridge University Press.
Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106,* 261-300.
Salthouse, T. A.(1992). Why do adult age differences increase with task complexity? *Developmental Psychology, 28,* 905-918.
Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403-428.
Seidenberg, M. S., McClelland, J. L. (1989). A distributed, developmental model of word

recognition and naming. *Psychological Review, 96*, 523-568.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experiment Psychology: General, 125,* 4-27.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*, 250-264.

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

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Journal of the ACM, 9,* 36-45.

Yuill, N, Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology, 80,* 351-361.