Working Memory Capacity and Categorization: Individual Differences and Modeling

Stephan Lewandowsky

School of Psychology

University of Western Australia

in press, *JEP:LMC*

Stephan Lewandowsky

School of Psychology

University of Western Australia

Crawley, W.A. 6009, AUSTRALIA

lewan@psy.uwa.edu.au

URL: http://www.cogsciwa.com

Abstract

Working memory is crucial for many higher-level cognitive functions, ranging from mental arithmetic to reasoning and problem solving. Likewise, the ability to learn and categorize novel concepts forms an indispensable part of human cognition. However, very little is known about the relationship between working memory and categorization, and modeling in category learning has thus far been largely uninformed by knowledge about people's memory processes. This article reports a large study (N = 113) that related people's working memory capacity (WMC) to their category-learning performance using the 6 problem types of Shepard, Hovland, and Jenkins (1961). Structural equation modeling revealed a strong relationship between WMC and category learning, with a single latent variable accommodating performance on all 6 problems. A model of categorization (ALCOVE) was fit to the individual data and a single latent variable was sufficient to capture the variation among associative learning parameters across all problems. The data and modeling suggest that working memory mediates category learning across a broad range of tasks.

Working Memory Capacity and Categorization: Individual Differences and Modeling

Is this furry animal a Siamese or a Ragdoll cat? Is this skin spot a harmless mole or an early-stage melanoma? The human ability to categorize objects into classes of like entities is acknowledged to be "basic to all of our intellectual activities" (Estes, 1994, p. 4). The importance of categorization is reflected in the intensity and success of research during the past several decades. We now understand the cognitive processes underlying categorization in considerable depth and an array of computational models provide detailed quantitative accounts of many phenomena (e.g., Anderson & Betz, 2001; Ashby & Maddox, 1993; Kruschke, 1992; Nosofsky, 1986; Nosofsky, Palmeri, & McKinley, 1994).

What is the sum of 3, 6, and 9? Which cognitive scientist was cited verbatim in the previous paragraph? The capacity to manipulate symbols while simultaneously remembering information for brief periods is the domain of working memory (WM). Like categorization, WM is a core concept in cognition that has attracted much empirical attention. The importance of WM is underscored by the fact that working-memory capacity (WMC) shares around 50% of the variance with people's general fluid intelligence (Kane, Hambrick, & Conway, 2005) and is predictive of performance in numerous reasoning tasks.

In light of the shared importance of categorization and working memory to cognition, their empirical and theoretical relationship is of considerable interest. For example, given the obvious involvement of memory in categorization, how does WMC contribute to people's ability to learn new categories? Given that the importance of attention has been emphasized in categorization as well as in working memory research, is there a common conception of attention that underlies both?

Perhaps surprisingly, only a handful of studies have addressed the linkage between WMC and category learning, with sometimes conflicting outcomes (Blair et al., 2009;

DeCaro, Thomas, & Beilock, 2008; DeCaro, Carlson, Thomas, & Beilock, 2009; Erickson, 2008; Tharp & Pickering, 2009). The purpose of this article is to begin to redress this deficit by providing a systematic theoretical and empirical analysis of the relationship between WMC and performance in the classic categorization tasks pioneered by Shepard, Hovland, and Jenkins (1961). This article first analyzes two primary theoretical linkages between categorization and working memory—viz. the role of WM in long-term learning and the emphasis on attention shared by both arenas—and briefly reviews the existing data on this issue. I then present a study that relates WMC to performance in all Shepard problems and apply a theory of categorization (ALCOVE) to the results at the level of individual participants. The parameters that co-vary with people's WMC are identified, thus creating a quantitative theoretical link between working memory and categorization processes. Structural equation modeling (SEM) is used throughout to describe the correlational structure among all behavioral measures and model parameters. To foreshadow the principal conclusions, the data suggest that performance in all categorization problems is strongly and uniformly associated with WMC. Within ALCOVE, variation in WMC is captured by variation in the associative learning parameter, suggesting that individual differences in performance are best modeled by different speeds of learning rather than differences in attentional abilities or differences in the precision of exemplar memory.

Working Memory and Long-Term Learning

Learning, by definition, involves creation of a memory for later access. Accordingly, models of categorization have at least tacitly assumed the operation of a memory (e.g., in exemplar models such as the GCM; Nosofsky, 1986; Nosofsky & Johansen, 2000; as well as in rule-based models such as the GRT; Ashby & Gott, 1988; Ashby & Maddox, 1993, or SRM; Johansen & Kruschke, 2005), and some models have explicitly stipulated the existence of multiple memory systems (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & O'Brien, 2005).

It is less obvious, perhaps, why category learning—a long-term process—should involve short-term and working memory. Notwithstanding some uncertainty about the underlying mechanism, there is much evidence that memory over the short term is a determinant of long-term learning. For example, simple nonword repetition tasks are strongly predictive of later language-acquisition skills (Baddeley, Gathercole, & Papagno, 1998): A child's ability to repeat out loud a nonword such as "ninkum" reveals much about how likely they are to acquire, say, Russian later in life. When a processing component is added to such simple memory tasks, for example by presenting an interleaved sequence of processing steps (e.g., arithmetic operations) and memoranda (e.g., letters that are to be recalled immediately after the sequence), performance on such "complex-span" tasks predicts long-term memory abilities in a wide range of circumstances (e.g., Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009; Unsworth & Engle, 2007). Complex-span tasks are a popular tool to measure WMC, and much research in the working memory arena converges on the conclusion that WMC is positively associated with performance in any (long-term) memory-reliant task *unless* the memory component is exclusively driven by a sense of familiarity. For example, Oberauer (2005) showed that WMC is unrelated to performance on a recognition test that can be performed on the basis of familiarity alone (i.e., when old items must be differentiated from items never before seen), whereas WMC is strongly associated with recognition performance when it involves recollection (i.e., when to-be-rejected "new" items had been seen before but in a different context, thus requiring information about the specific "binding" between items and their context).

It is clear that familiarity alone is insufficient to support category learning: During training, all items typically are "old" after a few trials, and learning to associate them with different categories is a quintessential example of the binding processes known to be mediated by WMC (Oberauer, 2005, 2009). It follows that current theorizing in working memory would expect WMC to mediate category-learning in a very broad and general manner.

This expectation of broad-based mediation stands in contrast to a far more circumscribed role of WM in some recent theories of categorization. Ashby and colleagues (e.g., Ashby et al., 1998; Ashby & Maddox, 2005; Ashby & O'Brien, 2005) have suggested that different memory systems support different types of categorization tasks. For example, Ashby and O'Brien (2005) differentiated between four different types of categorization tasks, each purportedly supported by a different memory system. Three of those tasks are relevant here because they are instantiated in the present study. (The fourth one, prototype distortion, arguably has lesser diagnostic utility; Zaki, Nosofsky, Jessup, & Unverzagt, 2003.)

Unstructured tasks are those that (a) involve only a few exemplars and (b) cannot be solved by any means simpler than memorization of those exemplars. Those tasks ostensibly involve "episodic" memory, a long-term repository seen as distinct from the more short-term agency of WM (Ashby & O'Brien, 2005). Performance on unstructured categorization tasks would thus be mediated by WMC only indirectly, via the link between WMC and long-term episodic memory performance. *Rule-based tasks*, by contrast, can be "learned via logical reasoning" (Ashby & O'Brien, 2005, p. 85), although the scope of this presumed logical reasoning is often limited to category structures that can be classified on the basis of a single dimension (e.g., Maddox, Love, Glass, & Filoteo, 2008; Markman, Maddox, & Worthy, 2006).[1] Rule-based tasks explicitly involve WM; and hence WMC ought to mediate performance on those tasks—an expectation that has found some recent empirical support (DeCaro et al., 2008, 2009). Finally, *information-integration* tasks are those that are neither unstructured nor solvable by a verbalizable rule; thus, they are amenable to a simpler solution than memorization of exemplars but nonetheless defy verbalization. Instead, information-integration tasks require the integration of two or more aspects of the stimulus at a pre-decisional stage (e.g., Ashby & Ell, 2001) and are the domain of a "procedural" memory system.[2] Performance on information-integration tasks is explicitly thought *not* to be mediated by WM; indeed, because people are assumed to operate with a bias towards explicit hypothesis testing, which may overshadow the output of the procedural system, the engagement of WM may be harmful to performance on information-integration tasks. It follows that WMC may be negatively correlated with performance—a counter-intuitive expectation that has also found some recent empirical support (DeCaro et al., 2008).

In summary, current theory and data in working memory would expect WMC to mediate performance in all categorization tasks. By contrast, a theory of categorization that postulates different memory systems expects the association between WMC and category-learning performance to differ with the structure of the to-be-learned category. At first blush, there appears to be some evidence for the latter prediction: DeCaro et al. (2008) reported a positive correlation between WMC and rule-based performance, accompanied by a *negative* correlation with information-integration performance (for related results, see Kloos & Sloutsky, 2008; Minda, Desroches, & Church, 2008). However, a recent re-evaluation of those results (DeCaro et al., 2009; Tharp & Pickering, 2009) has identified this negative correlation as artifactual (more on that in the General Discussion). The issue is therefore unresolved, calling for new and robust evidence about how WMC mediates category learning.

Attention in Working Memory and Categorization

There has been much theoretical emphasis on executive attention as an explanatory construct underlying variation in WMC (e.g., Kane, Bleckley, Conway, & Engle, 2001;

Kane, Conway, Hambrick, & Engle, 2007; Kane, Poole, Tuholski, & Engle, 2006). Executive attention is defined as "... comprising those domain-general processes that keep stimulus and goal representations accessible ... under conditions of interference, distraction, and response competition" (Kane et al., 2006, p. 750). Support for the attentional view comes from a number of sources, including experimental manipulations involving long-term memory. For example, Kane and Engle (2000) showed that WMC-related differences in susceptibility to the buildup of proactive interference (PI) were eliminated by a secondary task. That is, whereas people with high WMC suffered less PI than people with low WMC in a control condition, the introduction of a secondary complex finger-tapping task equated the susceptibility to PI between high-WMC and low-WMC individuals. This result suggests that in the absence of a secondary task, high-WMC individuals could use their additional attentional resources to suppress PI.

In categorization, attention has also been central to theorizing. Unlike in WM, theorizing in categorization can rely on rigorous quantitative models that describe the processes by which attention is allocated (e.g., Erickson & Kruschke, 1998; Kruschke, 1992; Kruschke & Johansen, 1999). There are two distinct conceptions of attention in category learning: The first and most common is known as dimensional attention and refers to the process that adapts responding to the relevance of stimulus dimensions. For example, when differentiating between African and Asian elephants, observers ought to attend to the dimension "ear size" rather than other features such as skin color or the shape and color of tusks. The second conception of attention is known as representational attention and was introduced by Erickson and Kruschke (1998). Representational attention refers to the process by which specific stimuli are associated with different representational modules—for example, some stimuli might be classified on the basis of one type of rule (e.g., size might optimally differentiate between violins and violas) whereas other stimuli within the same broad category (string instruments) might be classified on the basis another rule (e.g., shape might distinguish electric from acoustic guitars better than size; Erickson & Kruschke, 2002). Representational attention has also been shown to be involved in the dynamic re-coordination of modules of partial knowledge, for example in response to a hint about a previously hidden feature of the stimulus space (Sewell & Lewandowsky, in press). At a theoretical level, representational attention is captured by a stimulus-specific "gating" mechanism within a modular theory of categorization (ATRIUM; Erickson & Kruschke, 1998), which adjudicates between competing partial information delivered by different modules of knowledge and helps select a response.

At first blush, the executive attention construct in WM research appears related primarily to the notion of representational attention—which involves a high-level response-selection mechanism—than dimensional attention—whose dependence on the perceptual nature of the stimuli (e.g., Nosofsky, 1986) identifies it as a rather more low-level process. Nonetheless, a potential link between dimensional attention and the executive-attention construct in WM can be identified: Although dimensional attention is most often described in positive terms, for example by noting that "…more strongly attended cues are multiplied by a larger factor in responding and learning" (Kruschke, Kappenman, & Hetrick, 2005, p. 830), any shift of attention *towards* a relevant dimension is inescapably tied to a shift *away* from irrelevant dimensions. Thus, allocation of dimensional attention can be equivalently characterized as a process to suppress irrelevant information. The notion of suppression, then, forms a potential conceptual link between dimensional attention in categorization and executive attention—e.g., via its role in suppressing PI; Kane and Engle (2000)—in working memory.

It follows that WMC may be associated with people's ability to allocate dimensional attention during category learning, and the present study seeks to explore that potential link. To date, only one small-scale study has examined this issue: Blair et al. (2009) correlated performance on a single complex-span task with an index of attentional allocation derived from eye-movement data. Blair et al. found a moderate but significant *negative* correlation between the two measures early (but not late) in learning, suggesting that low-span individuals can allocate attention more effectively than high-spans prior to learning of the category.

The present study

The present study sought to illuminate the relationship between working memory and category learning guided by the two theoretical linkages just discussed. Precise and reliable measurement of WMC is essential to this endeavor. Psychometric research has established that WMC can be characterized by a general factor that accounts for a large proportion of variance, together with lower-order factors capturing specific content domains and cognitive functions (Kyllonen & Christal, 1990; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Oberauer, Süß, Wilhelm, & Wittmann, 2003). Although a number of tasks have been shown to have high loadings on the general WMC factor (Oberauer et al., 2000), in practice WMC has often been assessed quite narrowly within a single paradigm, namely the complex-span task (e.g., reading span, operation span; cf. Conway et al., 2005). This focus on a single task has been characteristic of all research involving categorization and WMC to date (Blair et al., 2009; DeCaro et al., 2008, 2009; Erickson, 2008), which is potentially problematic because it implies that the measurement of WMC has been contaminated with variance that is specific to the chosen task rather than to the intended psychological construct. To reduce the contribution of task-specific variance, the present study measured WMC using a battery of four tasks that were selected on a priori theoretical grounds and are known to have desirable psychometric properties (for details, see Lewandowsky, Oberauer, Yang, & Ecker, 2010).

By the same token, reliable measurement of categorization performance should also involve a battery of different tasks. The present study thus examined performance in all 6 problem types pioneered in the classic study by Shepard et al. (1961). Figure 1 shows the logical structure of the 6 problems, known as Types I – VI. Performance on these problems has been considered in great detail and much is known about how people (and even subhuman species; Smith, Minda, & Washburn, 2004) learn those problems (e.g., Feldman, 2006; Lafond, Lacouture, & Mineau, 2007; Love, 2002; Minda et al., 2008; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Smith et al., 2004; Vigo, 2006). The Shepard problems thus present an ideal test bed for examinations of the role of WMC.

Moreover, within the earlier taxonomy of tasks and memory systems (e.g., Ashby & O'Brien, 2005), "rule-based" tasks are clear isomorphs of a Type I problem whereas "information-integration" tasks have been identified as equivalent to Type IV (e.g., DeCaro et al., 2008, 2009; Waldron & Ashby, 2001). Based on Ashby, Ell, and Waldron (2003), Type II problems can also be classified as rule-based. Finally, Type VI has been explicitly identified as "unstructured" (Ashby & O'Brien, 2005). Although the previous literature is largely mute on their assignment, Types III and V ought also to be considered "information-integration" tasks as they share the "rule-plus-exception" structure with Type IV (Nosofsky, Palmeri, & McKinley, 1994) and require attention—and hence integration—across all three stimulus dimensions. It follows that the present study can test the expectation of the multiple-memory systems view that WMC should be selectively related to performance on rule-based tasks (Type I and II; also Type VI via the indirect link between WM and "episodic" memory) but not on information-integration tasks (Type IV and likely also Types III and V).

Finally, there is consensus that the involvement of dimensional attention differs between the various Shepard tasks (Kruschke, 1992; Nosofsky, 1984). For example, whereas attention is focused on one dimension for Type I, it is spread evenly across all dimensions for Type VI (for details, see Nosofsky, 1984, Table 1). It follows that the Shepard tasks can also illuminate the link between dimensional attention and WMC.

The relatively large sample size combined with the large number of measured variables permitted the use of Structural Equation Modeling (SEM) to examine the correlational structure among all measured variables. SEM is acknowledged to provide (largely) error-free measurement of latent constructs and the relationship between them (for a review see Tomarken & Waller, 2005) that cannot be obtained by analysis of pairwise correlations.

Method

The experiment was spread over three 1-hour sessions scheduled at the subjects' convenience but at least a day apart. The first session was devoted to measurement of WMC whereas the remaining two sessions involved learning of the 6 problems (3 per session).

Participants

The participants were 123 members of the University of Western Australia campus community who completed the experiment for partial course credit or A$10 per session. Data from 10 subjects were discarded either due to technical faults or because they failed to complete all experimental sessions. The analyses are therefore based on the data from 113 participants (28 males, age range 15–42, mean age 21.5 years). For 6 participants, WMC measurement occurred several weeks before the category-learning sessions, and those participants also took part in an unrelated experiment on memory updating in the meantime.

Apparatus

All experimental sessions were controlled by a Matlab program designed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Subjects were tested individually in small booths and stimuli were displayed on a 19 inch TFT monitor that was located approximately 70 cm in front of subjects.

Working Memory Capacity Measures

The first experimental session lasted 40 minutes and involved measurement of each participant's WMC using the battery presented by Lewandowsky et al. (2010) which consists of a memory updating task (MU), an operation-span task (OS), a sentence-span task (SS), and a spatial short-term memory task (SSTM), administered in that order. Those tasks are described in detail in Lewandowsky et al. (2010) and are outlined only briefly here.

The MU task required participants to (a) store a series of digits in memory, (b) mentally update those digits based on a series of arithmetic operations, and (c) recall the updated digits. On each trial, three to five frames were presented containing a random digit. A varying number of successive arithmetic operations, (e.g., '+4' or '-3') were then presented in the frames, one at a time, until the final result had to be recalled. There were a total of 15 trials.

On each trial of the SS task, several sentences were presented whose meaningfulness had to be judged. All sentences contained between 8 and 11 words and were followed by a consonant for memorization. List lengths ranged from 3 to 7 consonants, and keyboard recall was immediate. The OS task differed from SS only in that the sentences were replaced by arithmetic equations (e.g,. $4 + 3 = 7$) that had to be judged for correctness. List lengths ranged from 4 to 8. In both span tasks, there were 3 trials at each list length for a total of 15 trials.

The SSTM task involved memorization of the spatial location of circles in a 10×10 grid. On each trial, a series of solid black circles was presented, one-by-one, in various grid locations. The grid was then briefly removed before it reappeared without any circles and participants used the mouse to indicate the memorized location of the dots in any order by clicking in the corresponding grid cells.

The order of trials, selection of stimuli, and so on were randomized for all tasks. To reduce "method variance," the same randomization was used for all participants.

Category Learning

Design and stimuli

The category-learning sessions were designed to strike a balance between the need to maximize experimental control by counterbalancing and the desire to reduce method variance. This was achieved by creating 12 unique "modular sequences" to which participants were randomly assigned (with N ranging from 7 to 12). All participants assigned to one modular sequence received exactly the same treatment; that is, the same sequence of problems, the same randomization of stimuli in the training sequence and the same assignment of visual features to conceptual stimulus dimensions.

Across modular sequences, those variables were counterbalanced or randomized. Thus, 6 of the modular sequences involved all possible permutations of the easy problems (Types I–III) in the first category-learning session, followed by all possible permutations of the difficult problems (Types IV–VI) in the second session. For the remaining 6 modular sequences, the difficult permutations were administered in the first session and the easier ones in the second session.

Each problem involved 8 stimuli composed of 3 binary dimensions that were instantiated by three visual features, namely shape (square vs. circle), color (unfilled vs. red), and size (small vs. large). Each modular sequence involved a unique random mapping for each problem between those visual features and the underlying conceptual dimensions.

Procedure

Each problem involved a maximum of 12 training blocks, with 16 trials per block. Each block involved two presentations of the 8 stimuli in a random order that was unique to the particular modular sequence. A problem was terminated early if participants made no errors during two consecutive blocks, in which case perfect performance was imputed for the remaining blocks for analysis. Participants were aware that training would cease once they had reached this accuracy criterion. A self-paced break was inserted after every 4 blocks. Problems within a session were likewise separated by a self-paced break and instructions that signaled the onset of a new problem.

On each trial, the stimulus was centrally presented and remained visible until participants pressed the "/" or "Z" key (arbitrarily mapped to category 1 and 0, respectively). Feedback (the centrally presented word "Correct" or "WRONG") followed each response for 2 *s*. Trials were separated by 1 *s*.

Results

WMC Measures

WM tasks were scored using partial credit scoring (cf. Conway et al., 2005). For instance, a subject who correctly remembered 5 out of 6 letters in a complex-span trial would score 5/6 on that trial, with the person's total score representing the mean of these partial scores across trials. Descriptive statistics for the WMC battery are shown in Table 1; they are consonant with those of three experiments reported by Lewandowsky et al. (2010).

The relatively low performance on the sentence-meaningfulness task (SS_{pt}) is not unexpected given the difficulty of the material used here; for further details, see Lewandowsky et al. (2010). The table also shows that memory performance across participants ranged from safely above floor to near ceiling. Examination of the underlying distributions (not shown) identified 1, 0, 2, and 3 individuals who performed above .95 on the WMU, OS, SS, and SSTM tasks, respectively. All observations were retained for analysis.

Category Learning

Practice effects

Initial inspection of performance revealed large practice effects for the more difficult problems (i.e., Types IV, V, and VI): The later in the session a problem was presented, the better was average performance. In confirmation, a 12 (modular sequence) × 6 (problem type) × 12 (block) between-within ANOVA yielded a significant interaction between modular sequence and problem type, $F(55, 505) = 3.67$, $MSe = .086$, $p < .0001, \eta_p^2 = .29$, as well as a significant over-arching interaction between all three variables, $F(605, 5555) = 1.67$,$MSe = .009$, $p < .0001, \eta_p^2 = .15$. (The expected effects of block and problem type were also obtained and are explored later.)

Following precedent (Wilhelm & Oberauer, 2006), those practice effects and other counterbalancing "noise" captured by the modular-sequence manipulation were statistically removed via dummy regression. Specifically, for each problem type, performance on all blocks was regressed onto 12 predictors that dummy-coded (1 or 0) the modular sequence of each subject. The residuals of that analysis were added to the mean for that problem across subjects, thus yielding a final set of transformed scores that (a) preserved all individual differences and mean performance for each problem across all possible orders while (b) removing practice effects and other method variance associated with the modular sequences from each problem. All remaining analyses were conducted on these transformed scores.

Learning

Figure 2 shows the average learning curves for all problem types. In line with most prior research, the figure uses proportion of errors in each block as the dependent variable. The slightly scalloped shape of the curves arises from the self-paced breaks after Blocks 4 and 8. The principal features of the Shepard problems were replicated here: The Type I problem was learned considerably faster than any of the others, Type VI was most difficult, and the remainder clustered in between. The Type II problem was learned marginally faster than Types III–V. Although there are some reports in the literature of a considerable advantage for Type II relative to Types III–V (Nosofsky, Gluck, et al., 1994), it is common for that difference to be slight (Smith et al., 2004) or even virtually absent (Love, 2002).

Table 2 shows the average proportion correct for each problem (averaged across blocks for each participant and then averaged across participants) together with their standard errors and other summary statistics. The obvious pattern in the table and the figure was confirmed by a 6 (problem type) × 12 (block) within-subjects ANOVA, which yielded a significant main effect of problem type, $F(5, 560) = 37.66$, $MSe = .11$, $p < .0001, \eta_p^2 = .25$, and of block, $F(11, 1232) = 370.06$, $MSe = .01$, $p < .0001, \eta_p^2 = .77$, as well as an interaction between both variables, $F(55, 6160) = 5.48$, $MSe = .01$, $p < .0001, \eta_p^2 = .05$.

Linking WMC and Category Learning

Structural equation modeling was used to investigate the relation between WMC and category learning. Separate measurement models for WMC and category learning were obtained that were subsequently combined into a single structural model.

Measurement model for WMC

Table 3 shows the pairwise correlations among the four WM tasks, and also their correlations with category learning, that formed the departure point for the SEM analysis. The analysis used a single latent variable (labeled *WMC* from here on) that predicted the four manifest variables. The error terms associated with *OS* and *SS* were allowed to be correlated to reflect the particular similarity of the two complex-span tasks (standardized estimate .26, $p \cong .05$). This model provided a reasonable fit to the data, with $\chi^2(1) = 4.44$, $p < .05$; comparative fit index, $CFI = .973$; root-mean-square error of approximation, $RMSEA = .175$; standardized root-mean-square residual, $SRMR = .033$. The standardized regression weights for the four manifest variables are shown in Table 1. With the exception of the larger RMSEA, these fit indices and regression weights (loadings) are comparable to those reported in previous applications of the same battery Ecker, Lewandowsky, Oberauer, and Chee (2010); Lewandowsky et al. (2010).

The fit can be improved ($CFI = 1$; $RMSEA = 0$; $SRMR = .016$) if the SSTM and SS error terms are allowed to correlate instead of SS and OS. Because there are stronger theoretical reasons (and some precedent with this battery; Lewandowsky et al., 2010) to expect the span tasks to be correlated, the first model appears preferable despite the larger RMSEA.

Measurement model for category learning

The measurement model for category learning was, by itself, of theoretical interest because it illuminates the relationship among the various tasks and might shed light on the putative involvement of multiple memory systems. Table 3 shows the pairwise correlations between proportions correct (PC) across the 6 problem types (and also their correlations with the WMC measures). It is apparent from the table that all tasks were highly and positively correlated with each other, although there is a suggestion that the correlations involving Type I were tending towards lower values than any of the others.

The first candidate measurement model included only one latent variable that predicted performance for all 6 problems. This model fit moderately well by a CFI (.945) and SRMR (.0466) criterion, although other fit indices were less convincing; $\chi^2(9) = 29.7$, $p < .002$, RMSEA= .132 (90% CI: .076 – .192). A potential alternative two-factor model, with Types I, II, and VI loading on one variable and Types III, IV, and V on the other, proved identifiable only when the correlation between the two factors was set to unity (i.e., when the two latent variables were collapsed into one); when the correlation was free to vary, its standardized estimate exceeded 1.0. Further exploration thus focused on improvement of the single-factor model, using a stringent ($\alpha = .01$) criterion for improved fit as recommend by Ullman (2001).

On the basis of modification indices, the single-factor model was improved in two steps ($\Delta\chi^2(1) > \chi^2(1)_{crit;\alpha=.01}$ in each case) by adding pairwise correlations between the error terms for Type IV and Type V, and Type V and Type VI, respectively. This model achieved an excellent fit, $\chi^2(7) = 8.8$, $p > .1$; CFI= .994; RMSEA= .048 (90% CI: .0 – .132); SRMR= .0268. The standardized regression weights (all $p < .0001$) for the six manifest variables were .87, .80, .67, .71, .67, and .65, for Types I–VI, respectively.

The fact that a single latent variable —albeit augmented by two pairwise correlations between the error terms for Types IV through VI— was sufficient to capture variation in performance across participants is noteworthy in light of the rather striking differences in performance and speed of learning between problems (see Table 2). It must be noted that this conclusion is not altered if other manifest variables for category learning are considered (e.g., estimated learning rates based on fits of descriptive functions).

Structural model

The complete structural model that related working-memory capacity to the single latent variable capturing category-learning performance provided a reasonably good fit, with $\chi^2(31) = 62.4$, $p < .0001$; CFI= .938; RMSEA= .095 (90% CI: .06 – .129); SRMR= .0599. Figure 3 displays this final model.

Of greatest interest is the strong link (correlation .59) between the two latent variables, *WMC* and *Learn*, which was highly significant ($p < .001$) and which attests to the fact that working memory was positively and uniformly related to category learning performance.[3]

Brief mention must be made of a common criticism of WMC as an explanatory construct, which appeals to unexamined variables such as motivation or willingness to expend effort that are assumed to explain variation in measures of WMC as well as in the criterion tasks. Three reasons speak against this possibility: First, WMC is associated with *many*, but definitely not all (e.g., Kane et al., 2006; Tuholski, Engle, & Baylis, 2001), cognitive activities, thus calling into question a general motivational account. Second, participants in the present study knew that they could exit the experiment more quickly, and without loss of remuneration, if they performed perfectly for two blocks, thus providing a strong incentive to do well. Finally, WMC has been found to be uncorrelated with effort expended on the basis of pupillometric measures that are known to be sensitive indicators of mental effort (Heitz, Schrock, Payne, & Engle, 2008). The present data are therefore best accepted as illuminating the link between WMC and category learning, and their full implications are discussed after the computational modeling has been presented.

Computational Modeling

Two models of categorization that are known to handle the Shepard result—ALCOVE (Kruschke, 1992) and the configural-cue model (e.g. Gluck, 1991;

Nosofsky, Gluck, et al., 1994)—were fit to each participant's data from all problem types. The fits of both models support the same substantive conclusions. Because ALCOVE fit the data far better than the configural-cue model, only the results involving ALCOVE are reported in detail here; the results for the configural-cue model are available from the author upon request.

The modeling served three purposes: First, to test the models' ability to capture individual variation. At present, it is unknown whether any existing model is capable of accounting for individual differences in category learning at a fine level of detail. Second, to identify which parameters captured those individual differences and how variation in parameter values correlates across problems. Identification of parameters would provide valuable insights into likely ways in which different individuals approach the task. Finally, I sought to relate the observed parameter variation to WMC, thus pointing to a link between working memory and category learning at a process level.

General Considerations: Parameter dependency

There has been much recent emphasis on fitting models to the data from individuals rather than to the group average (e.g., Cohen, Sanborn, & Shiffrin, 2008). In many instances, fitting of individuals is preferable because grouping can distort the form of the data and may obscure large differences between individuals' strategies (e.g., Yang & Lewandowsky, 2004), although it raises another problem known as parameter dependency (e.g., Li, Lewandowsky, & DeBrunner, 1996; Ratcliff & Tuerlinckx, 2002; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). This problem is particularly relevant when the parameter estimates are used in correlational analyses, and refers to trade-offs between parameters that prevent an unambiguous interpretation. This problem is easily illustrated by considering a linear regression: Suppose one generates numerous data sets that conform to the equation $y = b + ax + e$, where $e \sim N(0, 1)$. Although all data sets, bar the random error, should yield identical estimates for parameters a and b, in actual fact a negative correlation between those parameters is observed, which can be as large as $-.848$ (Ratcliff & Tuerlinckx, 2002).[4] It follows that when models are fitted to individual subjects, variations in parameters represent true individual variation *and* the consequences of (nuisance) parameter covariations (e.g., Ratcliff & Tuerlinckx, 2002; Schmiedek et al., 2007). Given that in most situations, the nature of those covariations is unknown (though see Li et al., 1996, and Schmiedek et al., 2007, for possible analyses), it is difficult to disentangle nuisance covariation from true individual variation.

One solution to the parameter-dependency problem is to obtain multiple measurements of the parameters using different tasks that represent a common construct. If those multiple measurements are followed by a latent-variable analysis, the nuisance covariations between parameter estimates can be statistically removed and the resultant latent variable(s) capture(s) the true variation of the parameter(s) across individuals (Ratcliff, Thapar, & McKoon, in press; Schmiedek et al., 2007).

This technique was applied here: The behavioral data just presented suggest that the 6 Shepard problems tap a single psychological construct. It follows that if a model is fit to those 6 problems, estimating the parameters anew for each problem and participant, nuisance correlations between parameters (if any) can be disentangled from true variation across individuals by latent-variable analysis. Note that this approach does not constrain the parameter values for all problems to uniformly co-vary; it is possible for multiple latent variables to be required to capture their variance. Likewise, the approach does not mandate the relationship between WMC and the parameters, if any, to be uniformly positive. Instead, it is theoretically possible for a subset of parameter values, captured by one latent variable, to be negatively correlated with WMC whereas another latent variable with a different subset of parameter values might be positively correlated.

ALCOVE

ALCOVE is an exemplar-based connectionist model of category learning (Kruschke, 1992). Each training exemplar is represented by a node in category space whose location corresponds to the dimensional feature values of the stimulus. During training, ALCOVE learns to adjust the connection strengths between exemplar nodes and the possible responses using standard network learning rules. ALCOVE also learns, by a similar error-driven learning rule, how much attention to pay to each dimension, which results in the stretching of highly diagnostic dimensions and the shrinking of others that are less relevant.

In addition to those learning processes, another important contributor to ALCOVE's performance is the nature of its exemplar memory. Specifically, when presented with a test item for categorization, ALCOVE responds on the basis of the similarity between that item and every previously encountered exemplar from each category—this similarity judgment is affected by the "sharpness" of the exemplar memory; that is, the extent to which an exemplar node's response generalizes to other, similar stimuli. ALCOVE therefore provides another avenue by which differences in WMC might be modeled; viz. by the precision with which exemplars are remembered.

ALCOVE architecture

A full description of ALCOVE can be found in Kruschke (1992). The following brief summary of its architecture mainly serves to identify its four main parameters.

When a stimulus is presented, each exemplar node is activated according to its similarity to the input. Thus, the activation of exemplar node j is given by:

$$a_j^{hid} = \exp(-c(\sum_i \gamma_i \, |h_{ji} - a_i^{in}|)), \tag{1}$$

where h_{ji} is the location in category space of exemplar node j along dimension i, a_i^{in} the value of dimension i in the input (i.e., in the presented stimulus), and γ_i is the corresponding dimensional attention weight. Of particular interest here is the "specificity" parameter c, which determines the steepness with which activation declines as a function of the distance between the exemplar node and the stimulus. If c is large, then only the stimulus corresponding to the exemplar itself (or those in close proximity) will activate the node, whereas if c is small, there will be greater generalization and a node will be activated even by stimuli at a greater distance. In the present context, it is noteworthy that larger values of c translate into more specific—and hence better—memory for the exemplars: In an exemplar-based model of memory, known as SIMPLE, that instantiates many of the same principles that are also embodied in ALCOVE, the specificity parameter maps directly into memory performance (Brown, Neath, & Chater, 2007). This relationship raises the possibility that WMC might likewise be related to the value of c in ALCOVE.

Each exemplar node is connected to all output nodes, which represent the possible categorization responses. Specifically, the connection between exemplar node j and output node k is referred to as w_{jk}, and output nodes are activated to the extent that they receive activation along all incoming weights from all exemplars:

$$a_k^{out} = \sum_j w_{kj}\, a_j^{hid}. \tag{2}$$

Once the output units have been activated, the stimulus is categorized using a variant of Luce's (1963) choice rule:

$$P(K) = \frac{\exp(\phi\, a_K^{out})}{\sum_k \exp(\phi\, a_k^{out})}, \tag{3}$$

where $P(K)$ refers to the probability of choosing category K and where ϕ is a free parameter that governs whether responding is deterministic (large ϕ) or probabilistic ($\phi \simeq 1$). Deterministic responding means that whichever category "wins" the similarity comparison is *always* chosen as a response, whereas if responding is probabilistic, that category is chosen with a probability proportional to its similarity advantage.

The final two parameters of ALCOVE, λ_A and λ_W, are the learning parameters for the adjustment of dimensional attention and the associative weights between exemplar nodes and output nodes, respectively. During training, the model computes an “error signal” that is based on the discrepancy between the current output and the target response, and the learning parameters determine the extent to which that error signal is used for weight update (for details, see Kruschke, 1992).

ALCOVE simulation methodology

ALCOVE was fit to each participant and each problem separately, yielding 6×113 different sets of parameter estimates for analysis. Maximum-likelihood estimates for the parameters were obtained using a Newton-type algorithm (Schnabel, Koontz, & Weiss, 1985) as instantiated in the *nlm* function in the R package (R Development Core Team, 2005; Version 2.8.0) with a binomial probability model. Specifically, maximization entailed the log likelihood:

$$\ln L = \sum_i d_i \ln(p_i) + (n_i - d_i) \ln(1 - p_i), \tag{4}$$

where p_i is the predicted probability correct across trials in block i ($i = 1, 2, \ldots, 12$), and d_i and n_i refer, respectively, to the number of observed correct responses and total responses per block (all $n_i = 16$).

The starting values for the parameters were set to the values reported by Kruschke (1992) for his demonstration involving the Shepard problems; viz. $c = 6.5$, $\phi = 2.0$, $\lambda_W = .03$, and $\lambda_A = .0033$. ALCOVE is known to be very sensitive to the particular sequence of stimuli encountered during training (Lewandowsky, 1995), and the model was therefore fit to each person's exact training sequence for each problem.

ALCOVE and individual differences in category learning

The overall predictions of ALCOVE, obtained by averaging predicted response probabilities across all individual fits for each problem, are shown in Figure 4. In replication of previous applications (e.g., Kruschke, 1992; Nosofsky, Gluck, et al., 1994), the model reproduced the major features of the data: Type I performance was best, Type VI was worst, and the others clustered in between with a slight advantage for Type II (especially towards the end of training) over the others. Notably, the predicted differences between problem types appeared smaller than the empirically observed differences.

To ascertain the model's ability to accommodate individual differences, Figure 5 plots predicted individual performance against observed performance. If the model captured performance perfectly, all points in each panel would fall on the diagonal. By and large, for all problems, most data points clustered around the diagonal, suggesting that many (but not all) individuals were well described by the theory. Figure 5 also shows that whenever ALCOVE badly failed to capture a person's performance, it involved over-predicting the performance of people who performed poorly (for Types II, V, and VI).

Table 4 summarizes the underlying parameter estimates. The table shows that the median estimates across subjects and problems are in the vicinity of those reported by Kruschke (1992), although the values of the attention-learning parameter were considerably larger here (note that Kruschke, 1992, did not estimate parameters but reported a demonstration simulation only). The mean values of c across problems are noteworthy: whereas the specificity remained more or less constant across Types I through V, it was noticeably larger for Type VI. This hints at the possibility that performance for that problem type was entirely "exemplar-driven"; that is, there was no generalization between stimuli and each exemplar node was sharply tuned and only associated to one response category.

Another noteworthy aspect of the data in Table 4 is the moderately large interquartile spread (especially for ϕ), which hints at a very skewed distribution of parameter values. This was confirmed by visual inspection of the distributions (not shown). To ameliorate the skew, c, λ_A, and ϕ were raised to the power .1, and λ_W was transformed by taking its cube root. The transformed distributions deviated considerably less from normality and were used for the SEM modeling.

ALCOVE and WMC

I first explored a number of possible measurement models for ALCOVE's four parameters, using each parameter's best-fitting estimates for a given problem as a separate manifest variable. As a first step, the 6 values of each parameter were used as manifest variables for a separate single-factor model; this revealed that for c, ϕ, and λ_A there was only a single loading that accounted for more than 10% of the variance in parameter values—the remaining 5 were $< 10\%$ in all instances. By contrast, the single-factor model for λ_W yielded squared multiple correlations for all manifest variables that were $\geq .10$; in consequence, the preferred measurement model involved only λ_W. (Because λ_A is of theoretical interest, a model including both learning parameters was also developed; because this alternative model leads to exactly the same substantive conclusions it is not reported in detail here but available from the author upon request.)

The preferred measurement model involving only λ_W fit very well and included a pairwise correlation between the error terms for λ_W^4 and λ_W^5 that mirrored the correlation in the measurement model for the data (adding a second analogous correlation between λ_W^5 and λ_W^6 did not further improve the fit); $\chi^2(8) = 10.2$, $p > .1$, CFI= .938, RMSEA= .049 (90 % CI: .0 – .128), SRMR= .0538. The standardized regression weights were .33, .31, .30, .43, .52, and .48 for λ_W^1 through λ_W^6, respectively (all $p < .05$).

This measurement model was combined with the earlier WMC measurement model to form the final preferred structural model which is shown in Figure 6. This model fit extremely well, $\chi^2(32) = 24.4$, $p > .1$, CFI= 1.0, RMSEA= 0. (90% CI: .0 – .043), SRMR= .046. Table 5 displays the underlying correlations between the parameter estimates for λ_W and the WMC tasks.

The SEM results for ALCOVE thus buttresses the conclusions suggested by the behavioral data: Individual differences in categorization performance mapped into different speeds of learning within the model, and the mapping was uniform for all problems. It is notable that the specificity parameter, c, played no systematic role in accounting for individual differences. This speaks against the possibility that differences in WMC are best modeled by differences in the precision of people's exemplar memory.

General Discussion

Summary of results

The results of the experiment and the modeling are readily summarized: (1) Individual variation in category-learning performance could be captured by a single latent variable for all 6 problems. (2) WMC was strongly associated with that latent variable, and thus mediated learning of all possible category structures that can be formed from three binary dimensions. (3) When the data were modeled at the individual level by ALCOVE, all systematic variation in parameters involved the learning parameter (λ_W), and this variation was (4) again captured by a single latent variable for all problems. (5) The latent variable for the learning parameter was strongly associated with WMC.

Relationship to previous findings

Modeling individual differences

There is at least one precedent for fitting computational models to the data from individual participants while also seeking to explain variation in parameter estimates via another psychological construct. Schmiedek et al. (2007) measured response times in a large number of choice tasks and fit the EZ-diffusion model (Wagenmakers, Maas, &

Grasman, 2007) to the data from each subject. Schmiedek et al. found that the drift-rate parameter (which determines the speed with which evidence for a decision is accumulated) was most strongly associated with WMC in an SEM solution. Most relevant in the present context is the use of SEM in that study, which permitted Schmiedek et al. to prevent known statistical dependencies among the EZ-diffusion model's parameters from contaminating their results.

In categorization, by contrast, interest in individual differences has thus far primarily focused on either enumerating and classifying different strategies (e.g., Johansen & Palmeri, 2002; Lewandowsky, Roberts, & Yang, 2006; Nosofsky, Palmeri, & McKinley, 1994; Yang & Lewandowsky, 2003) or on addressing how those differences might best be modeled at the level of subgroups of subjects (e.g., Lee & Webb, 2005; Navarro, Griffiths, Steyvers, & Lee, 2006; Yang & Lewandowsky, 2004). Although those precedents have drawn much-needed attention to the ability of models to explain "...both how people are cognitively the same *and* how they are different" (Lee & Webb, 2005, p. 620), those efforts have not examined parameter variation at the level of each individual and have not related the observed individual differences to other explanatory constructs.

WMC and categorization

The studies by DeCaro et al. (2008) and direct follow-ups, and those by Minda et al. (2008), Blair et al. (2009), and Erickson (2008) appear to form the sum total of the existing literature on the role of individual differences in working memory and categorization. The present data go beyond those precedents in several ways.

DeCaro et al. (2008) explored the predictions of the multiple-memory systems view (e.g., Ashby & O'Brien, 2005) by contrasting rule-based and information-integration tasks. DeCaro et al. reported a dissociation between WMC and categorization performance: In their study, WMC was positively associated with performance on a rule-based task (an isomorph of a Type I problem but with a fourth, irrelevant dimension) but *negatively* associated with performance on an information-integration task (a Type IV isomorph with an additional dimension), thus apparently confirming a rather counter-intuitive prediction of the multiple-memory-systems view.

However, this result has already undergone considerable re-evaluation: Tharp and Pickering (2009) suggested that the outcome may have reflected on inappropriate performance measure; namely the number of trials to criterion, where the criterion was defined as 8 consecutive correct responses. Tharp and Pickering showed that this criterion was unacceptably lax and may have spuriously created a negative correlation between WMC and information-integration performance. In confirmation, in a small-scale replication of their earlier study, DeCaro et al. (2009) showed that WMC was *positively* associated with performance in both rule-based and information-integration tasks when a suitably strict performance criterion (16 consecutive correct responses) was used.[5]

Minda et al. (2008) tackled the same theoretical questions but pursued a developmental approach: Children of various ages and adults were compared on their ability to learn four of the Shepard problems (Types I – IV). In apparent support of the notion that rule-based problems rely on WM whereas information-integration tasks do not, adults were found to outperform children on Type II and Type III but not on the information-integration Type IV. (Type I performance was also identical across all age groups but that comparison was marred by a clear ceiling effect and is thus of little relevance.) However, there are several reasons why this result should be interpreted with caution. First, children differ from adults in many ways other than development of their WM, and differences between age groups therefore only indirectly illuminate the role of WM. Second, inspection of the obtained learning curves reveals that all age groups hovered in the vicinity of chance performance on the Type IV problem and continued to do so throughout the 48 trials of training (Minda et al., 2008, their Figure 2D). This pattern departs from other known results involving the Type IV problem. For example, in the present study, the error rate after 48 trials was below .2 (compared to almost .5 for adults in the study of Minda et al.). The reasons for this anomaly are unclear but the lack of learning prevents a meaningful interpretation of Type IV performance in their study.[6]

There are two further studies that have failed to find an association between WMC and overall categorization performance. Erickson (2008) presented participants with stimuli from 4 different categories, comprised of two pairs of categories that were located in different quadrants of the two-dimensional category space. Categories within each pair were separated by two partial boundaries at varying orientations. Category spaces of that type are conducive to "partitioning" (Yang & Lewandowsky, 2003, 2004); that is, the creation of independent modules of partial knowledge that are then coordinated to drive responding. Erickson (2008) found that performance of those people who partitioned the category space was positively associated with WMC; the accuracy of people who did not partition the space was unrelated to WMC. Notably, WMC did not mediate performance even for those individuals who used one-dimensional rules to classify items, in deviation from previous results (e.g., DeCaro et al., 2008, 2009) and the present data. The reasons for this surprising finding are unclear; however, it must be noted that Erickson (2008) used only a single task to measure WMC (involving just 12 complex-span trials). The inevitable presence of measurement error and task-specific variance may have precluded detection of a correlation between WMC and category learning.

A similar comment applies to the study by Blair et al. (2009) mentioned at the outset, which also used only one task (Cronbach's $\alpha = .78$; Unsworth, Heitz, Schrock, & Engle, 2005), and which also failed to find a relationship between WMC and performance on a rule-based and on an information-integration task. Blair et al. (2009) additionally found that individuals with low WMC seemingly were better at allocating attention prior to learning of a task; however, an alternative explanation of the data is that the lag between allocating attention (as measured by eye-movements) and category learning (as measured by responding) was greater for low-WMC individuals than high-WMC individuals.

In summary, available precedents have presented a somewhat heterogeneous picture. Studies that used a minimal assay of WMC (i.e., a single complex-span task) have not always succeeded in detecting a positive association between WMC and categorization, especially when the sample size was also small. Studies that have used more than one WM task, however, have found a positive association between WMC and categorization, although the total selection of category structures observed to date has remained rather limited. Against the background of this fairly circumscribed extant literature, I next explore the implications of the present study.

Theoretical implications

Implications for views of categorization

There is a long empirical and theoretical history of differentiation between the various Shepard problems for one theoretical reason or another (e.g., Feldman, 2006; Lafond et al., 2007; Minda et al., 2008; Nosofsky, Palmeri, & McKinley, 1994; Vigo, 2006). The present data are no exception and present a challenge or guiding constraint for several theories.

Multiple-memory systems view. The present data run counter to the expectation of the multiple-systems view that WM should only be involved in "rule-based" learning but not in "information-integration" tasks. Contrary to that expectation, there was no hint in the present data of a differential involvement of WM across the different tasks. The present data thus appear to challenge one principal tenet of the multiple-memory systems view, namely the circumscribed involvement of WM in category learning. However, before this conclusion can be accepted, other studies must be considered that have examined the contribution of WM not via individual differences but via a secondary-task manipulation. There have been several reports that a secondary task, such as a concurrent Stroop task, selectively interferes with rule-based but not information-integration learning (Minda et al., 2008; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006, 2007). On the assumption that secondary tasks tie up WM, those findings seemingly corroborate the selective role of WM in rule-based learning but not in other forms of categorization. However, there are several reasons to suggest that this conclusion may be premature.

First, Waldron and Ashby (2001) used a performance measure (trials to criterion, where the criterion was 8 consecutive trials correct) that is now known to be problematic (DeCaro et al., 2009; Tharp & Pickering, 2009). Moreover, Nosofsky and Kruschke (2002) showed that the data of Waldron and Ashby (2001) are actually compatible with a single-system account. Finally, in a thorough re-analysis of the data of Zeithamova and Maddox (2006), Newell, Dunn, and Kalish (2010) showed that if analysis follows common practice (e.g., Maddox & Ing, 2005; Zeithamova & Maddox, 2007) and consideration is restricted to people who (a) learned the categories and (b) paid attention to the secondary task, then the selective interference effect disappears. In confirmation, Newell et al. (2010) presented two new experiments in which a secondary task affected both rule-based and information-integration performance equally (or indeed disrupted the information-integration task more). In further support, Newell, Lagnado, and Shanks (2007) reported secondary-task interference with another probabilistic categorization task that is said to rely on the procedural memory system. Finally, Minda et al. (2008, Experiment 2) examined selective secondary-task interference with 4 of the Shepard problems. Minda et al. claimed that the secondary task (articulation of irrelevant random numbers) affected rule-based performance (Type II) but not information-integration performance (Type IV). This conclusion can be questioned on several grounds: First, they observed no interference effect on Type I performance, notwithstanding the fact that this task is rule-based par excellence (DeCaro et al., 2008; Waldron & Ashby, 2001). (A ceiling effect may explain this peculiarity.) Second, their conclusion that Type IV performance was *not* affected by a secondary task appears imprudent in light of the facts that (a) the secondary task reduced performance by 11% overall; (b) secondary-task performance was always below control performance throughout the 20 blocks of training (sometimes considerably so), and (c) there were only 6 subjects in each condition. Thus, Minda et al.'s conclusion rests on accepting the null hypothesis on the basis of a very small sample and in the face of a numerically rather large and consistent effect.

In summary, the existing evidence for a selective effect of secondary tasks on rule-based responding does not withstand rigorous scrutiny. Secondary tasks can interfere with both rule-based and information-integration responding (Newell et al., 2007, 2010), quite in line with the present data. Working memory, then, seems to be an integral component of all forms of categorization, contrary to that particular tenet of the multiple-memory systems view.

Care must be taken to delineate the impact of the present data—and those just reviewed—on the multiple-memory systems view. There is much support for the view from dissociations not involving working memory (e.g., Ashby et al., 2003; Maddox, Filoteo, Hejl, & Ing, 2004; Maddox & Ing, 2005) and in particular from neuroscience (for a recent review in support of the multiple-memory systems view, see Poldrack & Foerde, 2008). In fact, the neuroscientific data have been interpreted as mandating the rejection of the single-system approach " . . . because it fails to successfully explain a large body of data from neuropsychological, neuroimaging, and animal neuroscience studies" (Poldrack & Foerde, 2008, p. 203). The impact of the present data is thus strictly limited to the selective role of WM within the multiple-systems view. Nonetheless, it must be noted that a broader—and more radical—re-evaluation of the multiple-systems view has been initiated by (Newell, Dunn, & Kalish, in press). Newell et al.'s re-evaluation includes a thorough analysis of the neuroscientific evidence; the robustness of the view in the face of this ongoing re-assessment thus remains to be ascertained.

Attention learning and ALCOVE. The 6 Shepard problems require radically different allocations of dimensional attention. For Type I, all attention must optimally focus on a single dimension; for Type II, attention must be divided equally between two dimensions while the remaining one is ignored; and the other problems all require an approximately equal distribution of attention across all three dimensions (for details, see Nosofsky, 1984, Table 1). The role of attention is underscored by the fact that when dimensional attention cannot be selectively allocated because the stimuli are integral, the rank ordering of difficulty among problem types is strikingly altered, with the Type II problem becoming more difficult than all others bar Type VI (Nosofsky & Palmeri, 1996).

Within ALCOVE, dimensional attention is *learned* rather than estimated from the data as in the GCM. Given a uniform distribution of attention (i.e., .33 on each dimension) at the outset, it follows that Types I and II necessarily require much attention learning, whereas Types III–VI need not. Against this background it is notable that the attention-learning parameter (λ_A) played a very minor role in capturing performance across individuals and problems; indeed, as already noted, it did not enter into a measurement model on its own. It follows that the individual differences in category learning were not captured by different speeds of attentional learning in ALCOVE and, by implication, that dimensional-attention learning is unrelated to WMC.[7]

Implications for theorizing in WM

Attentional views of WM. As noted at the outset, there is considerable support for an executive-attention view of WM (e.g., Kane et al., 2007). At first glance, the fact that the attention-learning parameter contributed little to ALCOVE's ability to handle the data appears at odds with the attention view; if attentional ability is an important determinant of individual differences in WMC, why was there no notable relationship between dimensional attention and WMC in the present study?

In response, it must be recalled that dimensional attention, such as required in the Shepard tasks, is only one manifestation of attention in category learning. Its other manifestation, namely the representational attention that permits the coordination of different modules of knowledge (e.g., Erickson & Kruschke, 1998; Sewell & Lewandowsky, in press; Yang & Lewandowsky, 2004), may therefore be more analogous to the executive attention notion in WM research than dimensional attention. Although an initial examination failed to find a link between representational attention in categorization and WMC (Erickson, 2008), for possible reasons cited earlier, recent work in the author's laboratory has shown that WMC is strongly related to people's ability to re-coordinate modules of partial knowledge (Sewell & Lewandowsky, 2010). Taken together with the present results, this suggests that whereas dimensional attention is likely unrelated to WMC, representational attention in categorization may well be isomorphic to the executive-attention concept favored by some theorists of working memory.

The role of complexity in WM and categorization. There has been considerable recent interest in the notion of algebraic (e.g., Boolean) complexity and how it relates to categorization (Feldman, 2006) and working memory (Halford, Cowan, & Andrews, 2007). Although there is some debate about the most appropriate way to measure complexity of a category-learning problem (e.g., Lafond et al., 2007; Vigo, 2006), to a first approximation the Boolean complexity of the Shepard problems is predictive of the difficulty of learning. For example, Feldman (2006) reported a correlation of .84 between complexity and difficulty of learning.

Halford and colleagues (e.g., Halford et al., 2007) argue that problem complexity affects performance owing to the fact that more complex problems require more (and often too many) problem components to be held in WM. At first glance, one might therefore expect WMC to be more relevant for the more complex problems (e.g., Type III and above) than for the easier problems. Problems that require few partial elements to be held in WM should not exceed even the capacity of low-WMC individuals, and performance should thus co-vary little with WMC, whereas more complex problems should only be solvable by individuals with greater WMC capacity, thus exhibiting more WMC-related variation. This idea that problem complexity should mediate the role of WMC had considerable appeal some time ago. For example, Carpenter, Just, and Shell (1990) presented a computational model of performance in the Raven Progressive Matrices Test, a favored assay of fluid ability, and showed that performance was a function of the presumed capacity of working memory, with (simulated) low-WMC individuals being able to solve simple problems but not harder ones, thus modeling the presumed increase in WMC-related variance with problem difficulty.

Intriguingly, subsequent empirical examinations of this theoretically-presumed mediating effect of difficulty have consistently failed. For example, Unsworth and Engle (2005) showed that the correlations between WMC and performance on the Raven remained identical across the first three quartiles of item difficulty (the final quartile escaped analysis because few subjects managed to complete those items). The correlation also remained invariant as a function of the presumed number of "memory tokens"; that is, the number of elements that Carpenter et al. (1990) hypothesized to require storage in WM (for further similar results, see, e.g., Salthouse, 1993; Verguts & De Boeck, 2002).

The present data are exactly in line with those precedents and support the conclusion that although problem complexity has dramatic effects on performance, and although individual differences in performance are strongly predicted by WMC, the influence of WMC is invariant across complexity and involves something other than the problems' demand for partial information to be held in memory. If not memory for partial information, what then are the mechanisms by which WMC influences performance in category learning?

Binding in WM. Oberauer and colleagues (e.g., Oberauer, Süß, Wilhelm, & Sander, 2007) have developed a sophisticated tripartite approach to working memory. According to their model, WM involves three concentric "layers" of increasingly accessible and active information: The first layer corresponds to the activated portion of long-term memory, the second is known as a "direct-access region," and the final, most highly active layer is a single item that is in the "focus of attention." WMC is thought to be associated with the size of the direct-access region; that is, the number of items that are available for immediate processing.

A crucial property of the direct-access region is that it temporarily binds together representations that are required for cognitive operations. For example, item representations may be bound to their temporal context, they may be bound to a spatial location, and they may be transformed before being bound to a new or different context (e.g., Ecker et al., 2010). The notion of binding is particularly relevant in the present context because long-term learning is thought to involve transfer of information from the direct-access region to long-term memory.[8] Specifically, all Shepard problems require the binding together of the three item dimensions and their association with a category. However, at present the exact role of short-term binding during long-term learning is far from clear; Oberauer's model is thus best considered as a pointer towards future development of a process model of categorization with a working-memory component—or equivalently, the development of a model of WM that can do categorization. At the moment, no such model exists, although it is at least possible to shine a light in that direction within ALCOVE.

Towards a process model of WM in category learning. In a final simulation, I replaced the individual estimated values of the learning parameters in ALCOVE with a "rehearsal buffer" notion to point towards development of a process model of WMC within a categorization context.

Specifically, individual variation in WMC was represented in this final simulation as follows: Scores on the four WM tasks were averaged for each subject, and the inter-individual range of those (z-transformed) averages was linearly re-scaled to the range 0–1. Each person's re-scaled score was multiplied by 192 (the total number of actual training trials for each problem), and that final number was taken to be the number of additional "rehearsals" that were randomly spread through the training sequence. Each "rehearsal" involved another learning opportunity for a stimulus in the simulation. Thus, the person with the highest WMC effectively received double the number of actual learning trials, and the person with the lowest WMC received only the actual 192 trials, with everyone else in between placed according to their re-scaled WMC. This simulation thus explicitly modeled working memory as the capability to "re-present" or "rehearse" a stimulus during presentation to a greater or lesser extent.

For this simulation, λ_W and λ_A were clamped at a value corresponding to the first quartile of the freely-estimated values (Table 4). The first quartile was chosen because it provided a reasonable floor from which the additional rehearsals associated with larger WMC could build up performance. Thus, $\lambda_W = .0156$ and $\lambda_A = 4.403 \times 10^{-6}$ for all problems, with the remaining parameters (ϕ and c) freely estimated as before.

The results of this simulation are shown in Figure 7, using the same format as the earlier ALCOVE results. The figure shows that ALCOVE can mirror the data quite well when the individual variation in learning rates is replaced by an—independently measured and WMC-related—"rehearsal" parameter. I do not proffer this model as a complete instantiation of working memory, but rather as a prototype sketch of the type of theoretical developments that are now called for. Oberauer (2009)'s binding framework provides one highly plausible platform within which to develop such a model; there is an implied isomorphism between the ability to bind components in WM and the ability to update associative links in long-term learning.

Other promising developments that may ultimately yield a unifying model of WM and categorization can be found in the particle filter model (Sanborn, Griffiths, & Navarro, 2006) and the multi-agent model of Mathy and colleagues (Bradmetz & Mathy, 2008; Mathy & Bradmetz, 2004). It remains to be seen if those models can provide a theoretical integration of WM and category learning. Whatever form such an integrated model may ultimately take, the present results clarify that the relationship between WMC and category learning is uniformly positive and operates at the level of associative learning rather than some other process such as dimensional attention or precision of exemplar memory.

References

Anderson, J., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin and Review*, *8*, 629-647.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442-481.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Science*, *5*, 204-210.

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory and Cognition*, *31*, 1114-1125.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33-53.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372-400.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149-178.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, *9*, 83-89.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158-173.

Blair, M. R., Chen, L., Meier, K., Wood, M. J., Watson, M. R., & Wong, U. (2009). The impact of category type and working memory span on attentional learning in categorization. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 3127–3132). Austin, TX:

Cognitive Science Society.

Bradmetz, J., & Mathy, F. (2008). Response times seen as decompression times in boolean concept use. *Psychological Research*, *72*, 211–234.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

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

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

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, *15*, 692–712.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769-786.

DeCaro, M. S., Carlson, K. D., Thomas, R. D., & Beilock, S. L. (2009). When and how less is more: reply to tharp and pickering. *Cognition*, *111*, 415–421.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, *107*, 284-294.

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 170–189.

Erickson, M. A. (2008). Executive attention and task switching in category learning: Evidence for stimulus-dependent representation. *Memory & Cognition*, *36*, 749–761.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107-140.

Erickson, M. A., & Kruschke, J. K. (2002). *Multiple representations in inductive category learning: Evidence of stimulus- and time-dependent representation.* (Unpublished manuscript.)

Estes, W. K. (1994). *Classification and cognition.* New York: Oxford University Press.

Feldman, J. (2006). An algebra of human concept learning. *Journal of Mathematical Psychology*, *50*, 339-368.

Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science*, *2*, 50-55.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: a new hypothesis. *Trends in Cognitive Sciences*, *11*, 236–242.

Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, *45*, 119–129.

Johansen, M. K., & Kruschke, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 1433-1458.

Johansen, M. K., & Palmeri, T. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, *45*, 482-553.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*, 169–183.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). Oxford: Oxford University Press.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 336-358.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66–71.

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of "executive attention". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 749–777.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, *37*, 374–382.

Kloos, H., & Sloutsky, V. M. (2008). Whats behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General*, *137*, 5272.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1083-1119.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 830-845.

Kyllonen, P., & Christal, R. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence*, *33*, 1-64.

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology*, *51*, 57–74.

Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, *12*, 605-621.

Lewandowsky, S. (1995). Base-rate neglect in ALCOVE: A critical reevaluation. *Psychological Review*, *102*, 185-191.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, *42*, 571–585.

Lewandowsky, S., Roberts, L., & Yang, L.-X. (2006). Knowledge partitioning in categorization: Boundary conditions. *Memory & Cognition*, *34*, 1676–1688.

Li, S.-C., Lewandowsky, S., & DeBrunner, V. E. (1996). Using parameter sensitivity and interdependence to predict model scope and falsifiability. *Journal of Experimental Psychology: General*, *125*, 360-369.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, *9*, 829-835.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 227-245.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 100-107.

Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition*, *108*, 578–589.

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, *17*, 944-948.

Mathy, F., & Bradmetz, J. (2004). A theory of the graceful complexification of concepts and their learnability. *Current Psychology of Cognition*, *22*, 41–82.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1518–1533.

Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology*, *50*, 101-122.

Newell, B. R., Dunn, J. C., & Kalish, M. L. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*, *38*, 563–581.

Newell, B. R., Dunn, J. C., & Kalish, M. L. (in press). System*s* of category learning: Fact or fantasy? *The Psychology of Learning & Motivation*, *54*.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review*, *14*, 505–511.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classiffication. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *115*, 39–61.

Nosofsky, R. M., Gluck, M., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*, 352-369.

Nosofsky, R. M., & Johansen, M. (2000). Exemplar-based accounts of “multiple-system” phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375-402.

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*, *9*, 169-174.

Nosofsky, R. M., Palmeri, T., & McKinley, S. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53-79.

Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review*, *3*, 222–226.

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, *134*, 368–387.

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 51, pp. 45–100).

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences*, *29*, 1017–1045.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold,

M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York: Oxford University Press.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*, 167-193.

Pelli, D. G. (1997). The video toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews*, *32*, 197–205.

R Development Core Team. (2005). R: A language and environment for statistical computing.

Ratcliff, R., Thapar, A., & McKoon, G. (in press). Individual differences, aging, and iq in two-choice tasks. *Cognitive Psychology*.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*, 438-81.

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology*, *84*, 171–199.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In *Proceedings of the 28th annual conference of the cognitive science society* (pp. 726–731). Austin, TX: Cognitive.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, 414–429.

Schnabel, R. B., Koontz, J. E., & Weiss, B. E. (1985). A modular system of algorithms for unconstrained minimization. *ACM Transactions on Mathematical Software*, *11*, 419–440.

Sewell, D. K., & Lewandowsky, S. (2010). *Attention and working memory capacity: Insights from blocking, highlighting, and knowledge restructuring.*

Sewell, D. K., & Lewandowsky, S. (in press). Restructuring partitioned knowledge: The role of recoordination in category learning. *Cognitive Psychology*.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, 1-42. (13, Whole No. 517)

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General*, *133*, 398–414.

Tharp, I. J., & Pickering, A. D. (2009). A note on DeCaro, Thomas, and Beilock (2008): Further data demonstrate complexities in the assessment of information–integration category learning. *Cognition*, *111*, 410–414.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*, *1*, 31–65.

Tuholski, S. W., Engle, R. W., & Baylis, G. C. (2001). Individual differences in working memory capacity and enumeration. *Memory and Cognition*, *29*, 484-492.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics (4th Ed.)* (pp. 653–771). Needham Heights, MA: Allyn & Bacon.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence*, *33*, 67-81.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104–132.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.

Verguts, T., & De Boeck, P. (2002). On the correlation between working memory capacity performance on intelligence tests. *Learning and Individual Differences*, *13*, 37–55.

Vigo, R. (2006). A note on the complexity of boolean concepts. *Journal of Mathematical Psychology*, *50*, 501–510.

Wagenmakers, E.-J., Maas, H. L. J. van der, & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3-22.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*, 168-176.

Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus-response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology*, *18*, 18-50.

Yang, L.-X., & Lewandowsky, S. (2003). Context-gated knowledge partitioning in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 663-679.

Yang, L.-X., & Lewandowsky, S. (2004). Knowledge partitioning in categorization: Constraints on exemplar models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1045-1064.

Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unverzagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society*, *9*, 394–406.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*, 387–398.

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition*, *35*, 1380–1398.

Author Note

Preparation of this paper was facilitated by a Discovery Grant from the Australian Research Council to the author and Gilles Gignac and an Australian Professorial Fellowship to the author. I thank Daniel Little and David Sewell for contributions to the programming of ALCOVE and I thank Gilles Gignac for contributions to the structural-equation modeling. I also thank Ullrich Ecker, Brett Hayes, and Klaus Oberauer for comments on an earlier draft of this article. Address correspondence to the first author at the School of Psychology, University of Western Australia, Crawley, W.A. 6009, Australia. Electronic mail may be sent to lewan@psy.uwa.edu.au. Personal web page: http://www.cogsciwa.com.

Footnotes

[1] Ashby et al. (1998) allowed for the possibility that even two-dimensional problems may be “rule-based,” provided both dimensions are expressed and measured in the same units (e.g., height and width of a square). This was relaxed further by Ashby et al. (2003), who noted that conjunction rules may render a multi-dimensional task rule-based even when stimulus dimensions—as in the present experiment—are not strictly commensurate.

[2] The labeling (and number) of the memory systems that are thought to underlie the various tasks has undergone some evolution. Initially, only two memory systems, known as explicit and procedural, were thought to drive responding in rule-based and information-integration tasks, respectively (Ashby et al., 1998). The classification and terminology used here relies on more recent writing (Ashby & O’Brien, 2005).

[3] For completeness’ sake, a two-factor structural model was fit with one latent variable representing Types I, II, and VI, and the other variable representing the remaining types. This model fit worse than the single-factor model, $\chi^2(31) = 73.0$, $p < .0001$; CFI= .917; RMSEA= .110 (90% CI: .077 – .143); SRMR= .0563. The estimated correlation between the two latent categorization variables was 1.0, and each was highly correlated with *WMC*, confirming the conclusions implied by the preferred single-factor model.

[4] The explanation is simple: if error is introduced into a single data point, then unless that point is at the centroid of the bivariate cloud, it will necessarily raise the intercept and lower the slope (or vice versa). By extension, even if all data points are disturbed, only a few will have extreme errors, but they will suffice to introduce a negative correlation between slope and intercept.

[5] In further replication of their earlier study, DeCaro et al. (2009) also reported an apparent (and transient) advantage for low-WMC participants on the information-integration task when the laxer performance criterion was used. However, this

result must be approached with some caution because of the small sample size (low-WMC $N = 14$).

[6] Minda et al. (2008) argued that Type III ought to be considered "rule-based" in the sense of being learned by the explicit system. However, their stated rationale can be rephrased to apply also to Type IV, which is uniformly considered to represent an information-integration task by other authors (DeCaro et al., 2008; Minda et al., 2008; Waldron & Ashby, 2001), and is thus unconvincing.

[7] At first glance, the negligible attentional involvement in the present modeling is at odds with previous attempts to model WM in ALCOVE. Nosofsky and Kruschke (2002) applied ALCOVE to the data of Waldron and Ashby (2001) and found that the selective impairment of rule-based learning could be modeled by a presumed disruption of attention-learning by the secondary task. Because rule-based learning, for the reasons just noted in connection with the Type I and Type II problems, requires re-distribution of attention, this manipulation selectively impaired rule-based but not information-integration learning. In light of the recent re-evaluation of the data of Waldron and Ashby (2001) by Newell et al. (2010), Nosofsky and Kruschke (2002) may have targeted a result that actually is in little need of explaining. A uniform impairment of all categorization problems by a secondary task (as reported by Newell et al., 2010), by contrast, is likely to require a reduction in the associative learning parameter in ALCOVE, quite in line with the modeling reported here.

[8] Oberauer (2009) elaborates on the presumed transfer process and additionally notes the importance of transformation of the information, from a temporary relational format to a unitized or "chunked" structure.

Table 1

Performance on the Working Memory Tasks

Measure	WMU	OS	OS_{pt}	SS	SS_{pt}	SSTM
Mean	.59 (.017)	.70 (.013)	.90 (.009)	.68 (.017)	.63 (.009)	.85 (.006)
Minimum	.21	.31	.44	.18	.11	.58
Maximum	.99	.92	1.00	.97	.77	.97
Kurtosis	2.13	3.20	9.96	3.14	9.12	5.67
Skewness	−0.053	−0.64	−2.44	−0.79	−1.75	−1.15
SEM weights	.83	.69		.55		.63

Legend. MU, Memory Updating; OS, Operation Span; SS, Sentence Span; $_{pt}$ denotes processing tasks; SSTM, Spatial Short-Term Memory.

Note. Standard errors in parentheses. SEM weights refer to standardized regression weights (also known as loadings) for the four tasks in the WMC measurement model.

Table 2

Proportion correct (PC) for all problem types

Measure	Type I	Type II	Type III	Type IV	Type V	Type VI
Mean PC (SE)	.96 (.01)	.88 (.01)	.88 (.01)	.86 (.01)	.84 (.01)	.80 (.01)
Minimum	.50	.45	.47	.52	.45	.45
Maximum	1.07	1.07	1.01	1.08	1.06	1.09
Kurtosis	22.21	5.93	5.47	4.79	4.28	2.36
Skewness	−3.99	−1.68	−1.54	−1.24	−1.15	−.62

Note. Standard errors (SE) in parentheses. Note that these are proportions correct after removal of order effects; hence individual observations (i.e., maximum) can exceed 1.0, unlike the means which are unaffected by removal of order effects.

Table 3

Correlations between all WMC tasks and category learning performance

	WMC tasks				Category learning tasks					
	WMU	OS	SS	SSTM	Type I	Type II	Type III	Type IV	Type V	Type VI
WMU	—									
OS	.559	—								
SS	.488	.536	—							
SSTM	.525	.474	.257	—						
Type I	.315	.340	.228	.374	—					
Type II	.292	.294	.190	.325	.529	—				
Type III	.282	.417	.188	.338	.585	.704	—			
Type IV	.329	.268	.109	.238	.536	.517	.613	—		
Type V	.394	.309	.261	.252	.395	.558	.570	.675	—	
Type VI	.499	.463	.156	.436	.396	.527	.556	.519	.596	—

Legend. MU, Memory Updating; OS, Operation Span; SS, Sentence Span; SSTM, Spatial Short-Term Memory.

Table 4

Summary of best-fitting parameter estimates for ALCOVE fitted to the data from each subject and problem type separately.

		Parameter			
Problem type		c	λ_W	λ_A	ϕ
Overall	Mean	7.81	.30	7.04	33.75
	Q1	6.40	.02	.00	2.00
	Median	6.50	.07	.08	2.58
	Q3	6.53	.53	.47	14.35
I	Mean	6.93	.48	1.94	12.29
II	Mean	7.67	.28	2.87	20.31
III	Mean	6.54	.29	7.49	22.89
IV	Mean	6.20	.26	1.35	10.14
V	Mean	7.18	.26	2.58	103.40
VI	Mean	12.36	.22	25.99	33.48

Q1; first quartile, Q3; third quartile

Table 5

Correlations between all WMC tasks and ALCOVE parameter estimates

	WMC tasks				ALCOVE parameter estimates					
	WMU	OS	SS	SSTM	λ_W^1	λ_W^2	λ_W^3	λ_W^4	λ_W^5	λ_W^6
WMU	—									
OS	.559	—								
SS	.488	.536	—							
SSTM	.525	.474	.257	—						
λ_W^1	.218	.273	.197	.297	—					
λ_W^2	.215	.161	.071	.238	.182	—				
λ_W^3	.204	.148	.172	.171	.246	.290	—			
λ_W^4	.187	.129	.126	.178	.206	.101	.087	—		
λ_W^5	.269	.228	.192	.147	.238	.186	-.021	.289	—	
λ_W^6	.149	.107	.122	.114	.123	.214	.083	.153	.113	—

Legend. MU, Memory Updating; OS, Operation Span; SS, Sentence Span; SSTM, Spatial Short-Term Memory.

Figure Captions

Figure 1. The 6 problem types introduced by Shepard, Hovland, and Jenkins (1961). For each problem type, the panel shows items in one category by filled circles and those in the other category by open circles. Each stimulus is defined by three dimensions that correspond to the edges of the cube in each panel. The bottom panel (g) identifies the stimuli by number; all problem types use the same labeling of stimuli.

Figure 2. Learning curves for all problem types. Each observation represents the mean error rate within a block averaged across all participants.

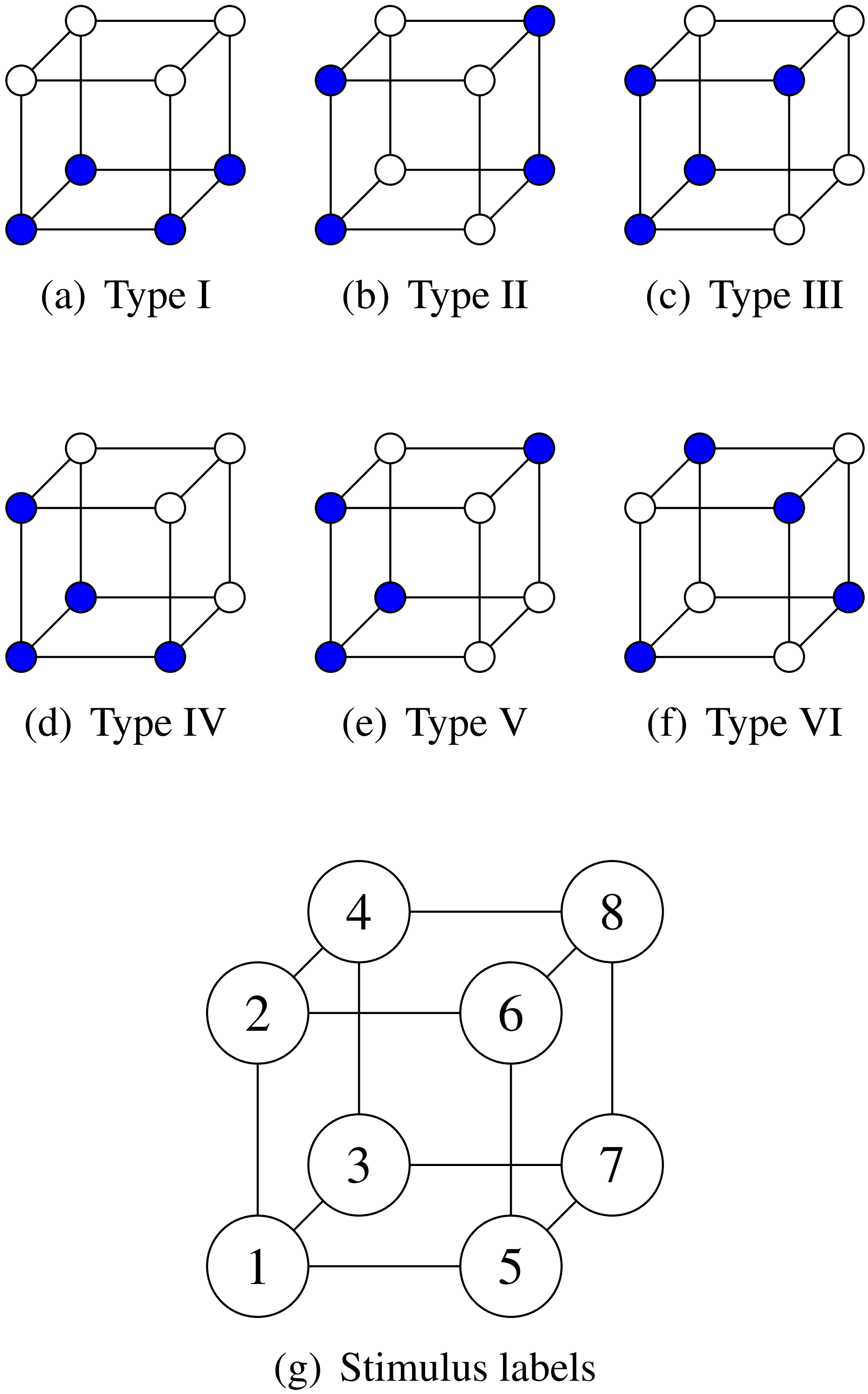
Figure 3. Structural model relating working-memory capacity (*WMC*) to category learning performance (*Learn*). Manifest variables refer to working-memory tasks (top) and the average proportion correct across training blocks for the 6 problem types (bottom). See text for further explanation.

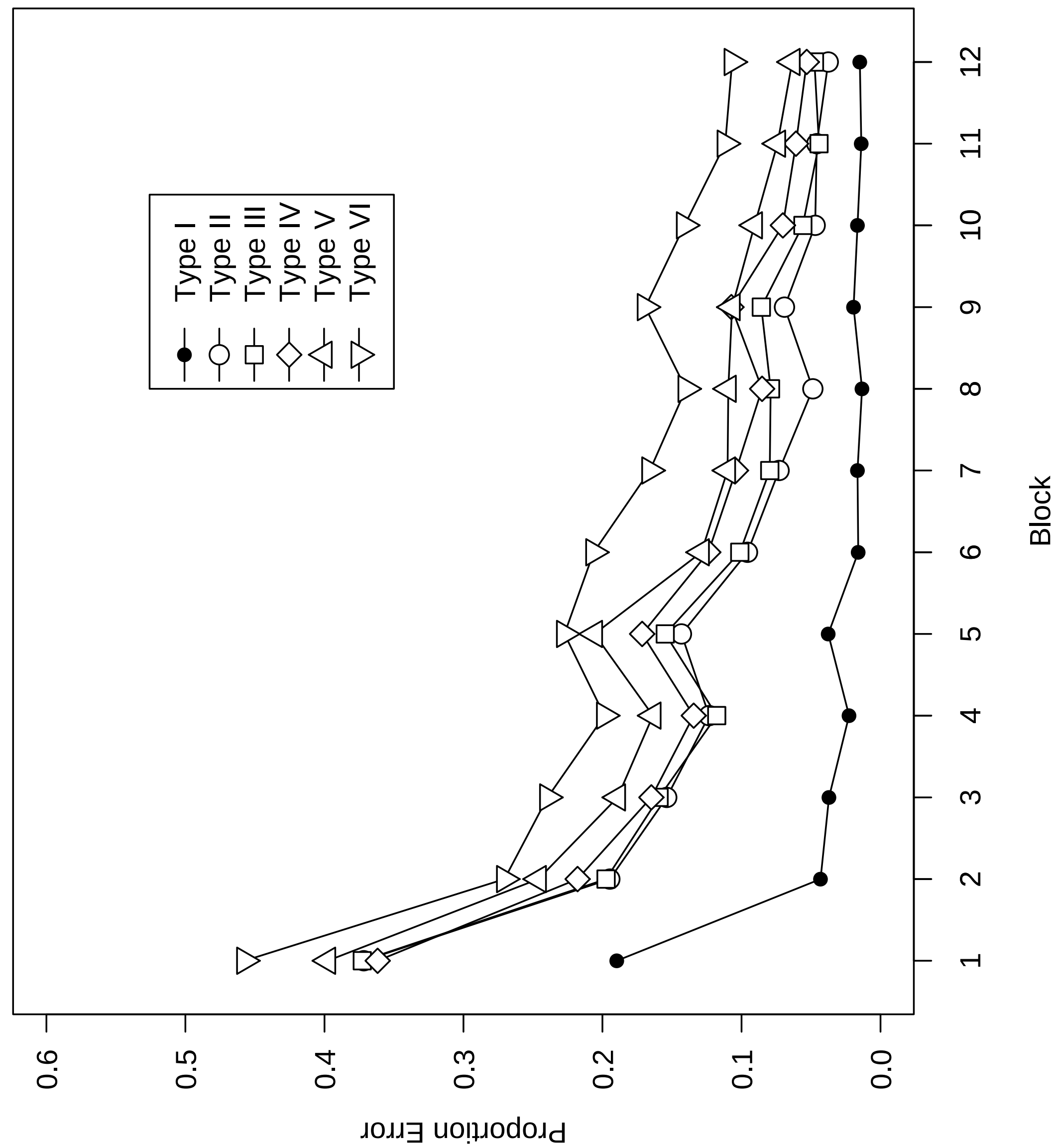
Figure 4. Average predictions of ALCOVE, based on separate parameter-estimates for each participant and problem. See text for details.

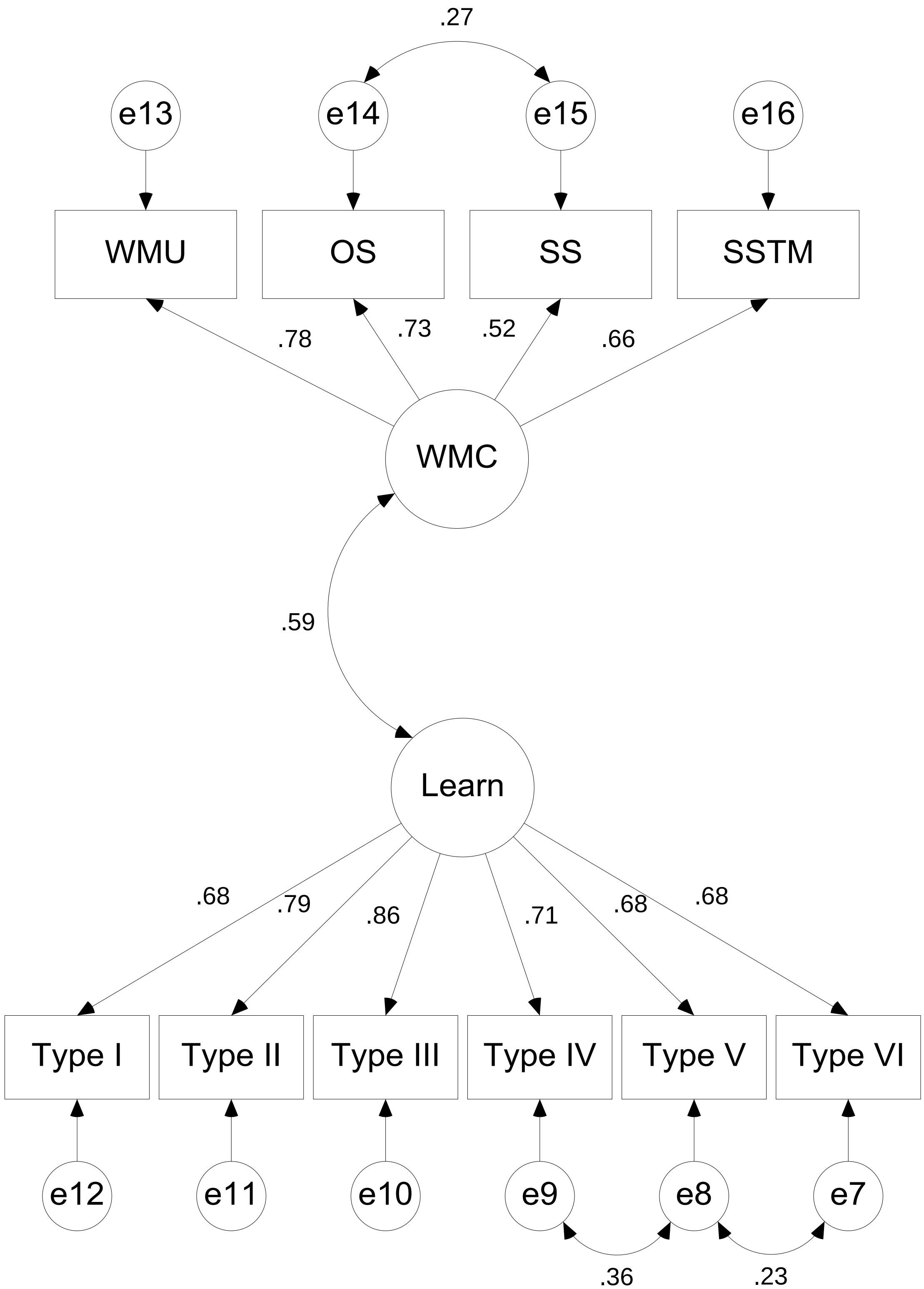
Figure 5. Individual performance predicted by ALCOVE (on the ordinate) and observed individual performance (on the abscissa) for problem Types I through VI. Each panel summarizes data and predictions for 113 subjects. The values of r^2 for the 6 problems are .86, .70, .88, .77, .60, and .76, respectively.

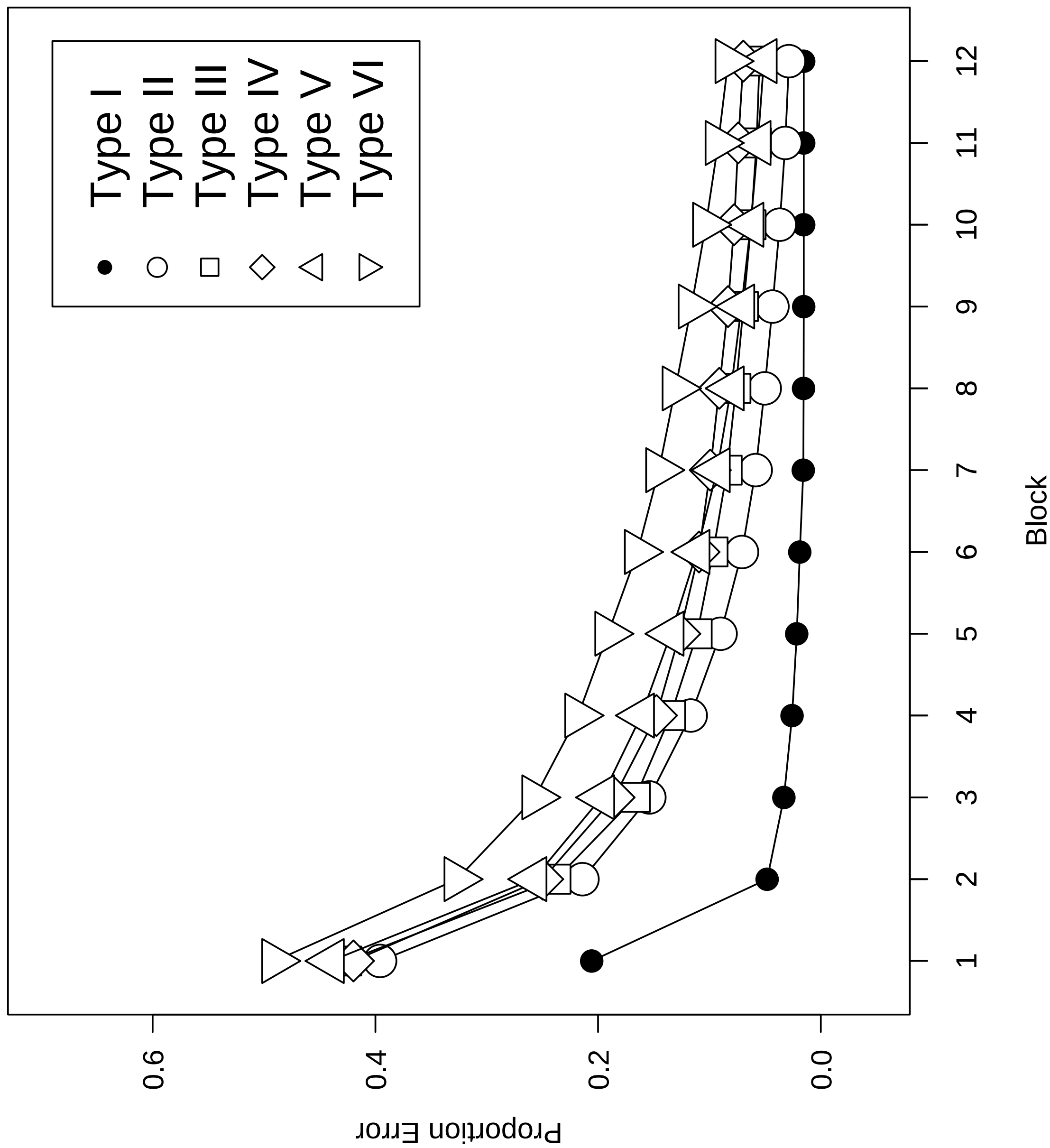
Figure 6. Final structural model relating the weight-learning parameter in ALCOVE (latent variable labeled *Learning parameter*) to the observed measurement of working memory (*WMC*). All weights and correlations are standardized, and manifest variables lambW1– lambW6 refer to $\lambda_W^1 - \lambda_W^6$, respectively.

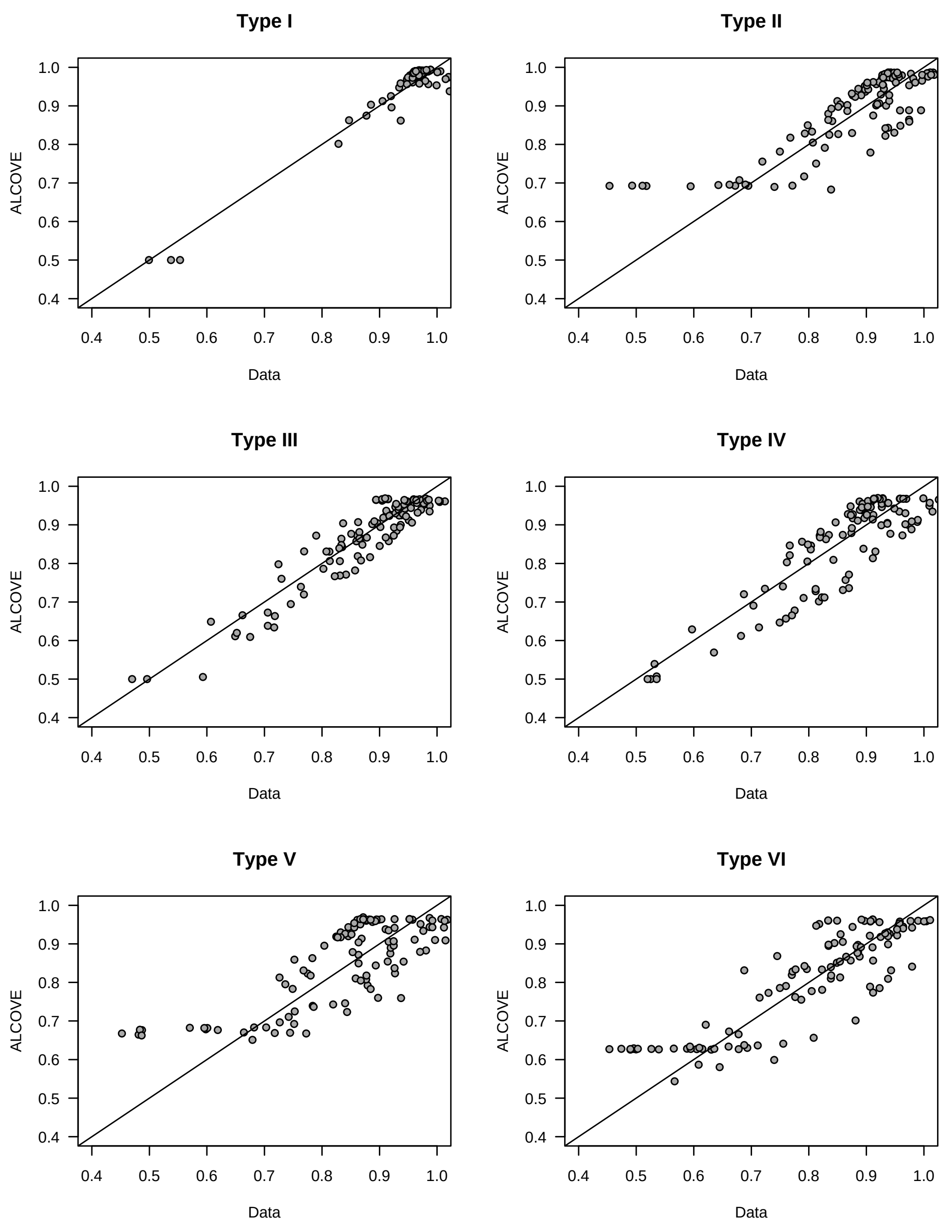
Figure 7. Observed category learning performance (abscissa) and performance predicted by the "rehearsal" version of ALCOVE (ordinate). Panels 1–6 show results for problem Type I through VI, respectively. Each panel summarizes data and predictions for 113 subjects.

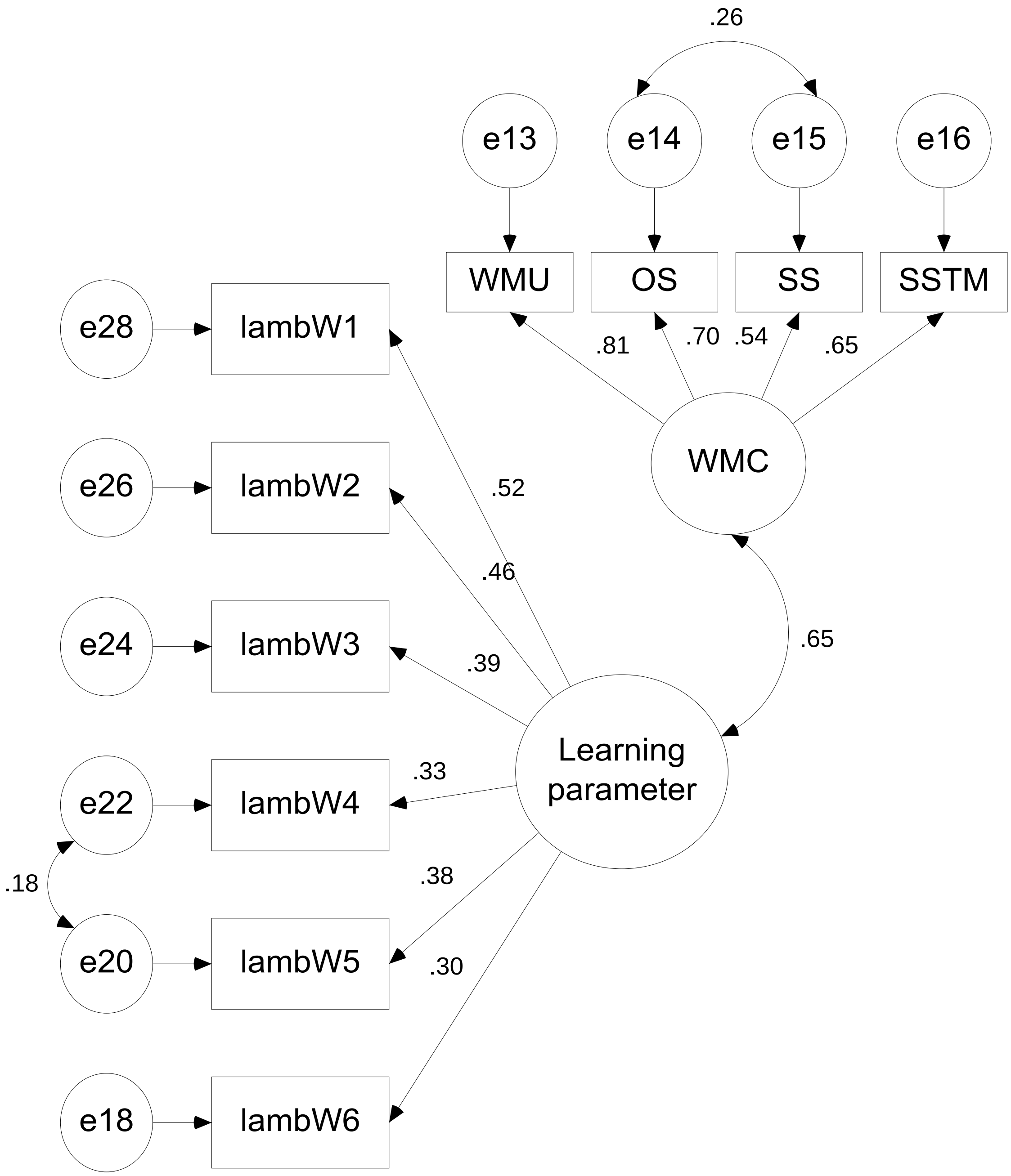

(a) Type I
(b) Type II
(c) Type III
(d) Type IV
(e) Type V
(f) Type VI
4
8
2
6
3
7
1
5
(g) Stimulus labels

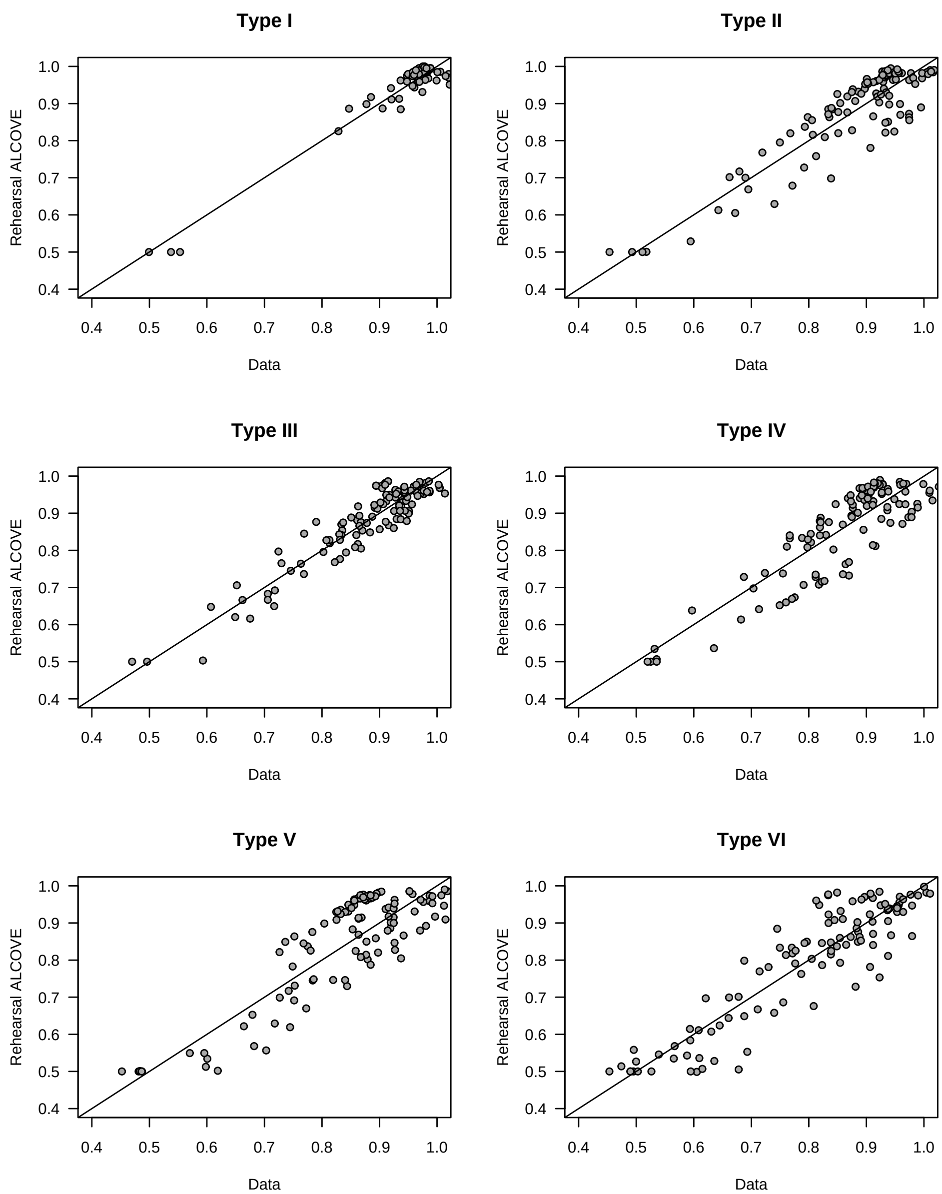

Type I
Type II
Type III
Type IV
Type V
Type VI
Block
Proportion Error
1
2
3
4
5
6
7
8
9
10
11
12
0.0
0.1
0.2
0.3
0.4
0.5
0.6


.27
e13
e14
e15
e16
WMU
OS
SS
SSTM
.78
.73
.52
.66
WMC
.59
Learn
.68
.79
.86
.71
.68
.68
Type I
Type II
Type III
Type IV
Type V
Type VI
e12
e11
e10
e9
e8
e7
.36
.23


Type I
Type II
Type III
Type IV
Type V
Type VI
Proportion Error
0.0
0.2
0.4
0.6
Block
1
2
3
4
5
6
7
8
9
10
11
12


Type I
Type II
Type III
Type IV
Type V
Type VI
ALCOVE
Data
0.4
0.5
0.6
0.7
0.8
0.9
1.0


.26
e13
e14
e15
e16
WMU
OS
SS
SSTM
.81
.70
.54
.65
WMC
e28
lambW1
e26
lambW2
e24
lambW3
e22
lambW4
e20
lambW5
e18
lambW6
.52
.46
.39
.33
.38
.30
.65
.18
Learning
parameter


Type I
Type II
Type III
Type IV
Type V
Type VI
Rehearsal ALCOVE
Data
0.4
0.5
0.6
0.7
0.8
0.9
1.0