Running head: ATTENTION AND WORKING MEMORY CAPACITY

Attention and Working Memory Capacity: Insights from Blocking, Highlighting, and Knowledge Restructuring

David K. Sewell

The University of Melbourne

Stephan Lewandowsky

The University of Western Australia

David K. Sewell

Psychological Sciences

The University of Melbourne

Melbourne, VIC 3010, AUSTRALIA

dsewell@unimelb.edu.au

Tel: +61 (03) 9344 8156

Fax: +61 (03) 9347 6618

Abstract

The concept of attention is central to theorizing in learning as well as in working memory. However, research to date has yet to establish how attention as construed in one domain maps onto the other. We investigate two manifestations of attention in category- and cue-learning to examine whether they might provide common ground between learning and working memory. Experiment 1 examined blocking and highlighting effects in an associative learning paradigm, which are widely thought to be attentionally mediated. No relationship between attentional performance indicators and working memory capacity (WMC) was observed, despite the fact that WMC was strongly associated with overall learning performance. Experiment 2 used a knowledge restructuring paradigm, which is known to require recoordination of partial category knowledge using representational attention. We found that the extent to which people successfully recoordinated their knowledge was related to WMC. The results illustrate a link between WMC and representational—but not dimensional—attention in category learning.

Attention and Working Memory Capacity: Insights from Blocking, Highlighting, and Knowledge Restructuring

Performance in many aspects of day to day life would be extremely difficult without the ability to hold multiple objects in mind at once. For example, reading a novel involves semantic processing of individual words in addition to the ability to relate those words to other aspects of the text such as plot and character development. The active and selective maintenance of different cognitive objects is the domain of working memory. Accordingly, the construct of working memory has considerable scope and serves a number of functions concerning the selection, maintenance, and manipulation of goal-relevant information (Oberauer, Süß, Schulze, Wilhelm, & Wittman, 2000; Oberauer, Süß, Wilhelm, & Wittman, 2003).

Much theorizing has linked people's working memory capacity (WMC) to the notion of attention. For example, some theorists have interpreted WMC as the number of cognitive objects that can either be simultaneously apprehended in the so-called "focus" of attention (e.g., Cowan, 2001, 2005), or immediately entered into the focus of attention (Oberauer, 2009). Other theorists have emphasized the supervisory function of the working memory system, identifying WMC with control over *executive attention* (e.g., Engle, 2002; Engle & Kane, 2004; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane, Conway, Hambrick, & Engle, 2007), which expresses itself in the ability to simultaneously maintain *relevant* task information, while also suppressing *irrelevant* information. In support, WMC has been found to be associated with resistance to interference in attentional-control situations without any memory involvement: Kane et al. (2001) showed that in a simple "antisaccade" task, in which a cue flashed on one side of the screen requires an orienting response to the opposite side, WMC was a predictor of performance. Similarly Kane and Engle (2003) showed that

WMC mediated performance on the classic Stroop task (i.e., naming the color ink of an incongruent word; e.g., “BLUE” presented in red ink).

The close theoretical relationship between attention and WMC invites the question of how WMC relates to other cognitive domains that ascribe a key role to attention. Under an executive-attention view, WMC might be expected to correlate with performance in learning tasks that involve deployment of selective attention. This article thus explores the relationship between WMC and two distinct manifestations of attention in category and associative learning; namely, dimensional attention on the one hand and representational attention on the other.

Attention and Category Learning

The idea that attention plays a role in category learning has a long theoretical history. Indeed, attention is a staple construct in contemporary models of categorization (see Kruschke, 2008, for a recent review). In categorization, attention is most often thought of in terms of a weighting of inputs to some stage of processing. The notion that stimulus information is selectively weighted has roots in classical views on attention (e.g., Kahneman, 1973; Neisser, 1967; Treisman, 1969) and remains at the core of modern theories of attention (e.g., Logan, 2002, 2004). To illustrate, a schematic overview of the stages of processing involved in categorization is shown in Figure 1. The figure shows a number of different processing stages and identifies two points at which attentional mechanisms have been thought to operate. Perceptual processes provide raw input that is subsequently weighted by *dimensional attention*, such that information along attended dimensions is accentuated, whereas information along unattended dimensions is attenuated (Nosofsky, 1986). Traditionally, only a single type of category representation is assumed to be available, in which case that representation is activated by the dimensionally weighted stimulus input, usually on the basis of similarity.

More recent models have investigated the possibility that multiple types of category representations might be available—for example, people may have access to a rule representation in addition to an exemplar-based representation to handle exceptions to the rule (e.g., Erickson & Kruschke, 1998). When multiple category representations are available, *representational attention* determines which one is selected, after which the selected category representation is activated by the dimensionally weighted stimulus input.[1] Activation of the category representation then drives a decision mechanism, resulting in an overt category response.

The attention-categorization link was first proposed by Shepard, Hovland, and Jenkins (1961), who examined the rates at which people learned to categorize stimuli defined along three binary-valued dimensions (e.g., large/small, red/green, triangle/square). Shepard et al. (1961) found that the rates at which people learned the 6 basic "problem types" that can be created from three binary dimensions corresponded to the number of relevant stimulus dimensions involved: The Type I problem (one relevant dimension) was learned the fastest, followed by Type II (two relevant dimensions), followed by the rest (three relevant dimensions). Moreover, learning rates were faster than could be predicted on the basis of stimulus generalization alone, with the discrepancy between theory and data being largest for the simpler problems (viz. Types I and II). Shepard et al. (1961) conjectured that selective attention to *relevant* stimulus dimensions might account for the discrepancy. Nosofsky (1984) provided theoretical support for this notion, showing that an extended version of Medin and Schaffer's (1978) context model with differentially-weighted stimulus dimensions produced a far better approximation to the data (see also Kruschke, 1992; Love, Medin, & Gureckis, 2004). The need for selective dimensional attention to quantitatively account for the data of Shepard et al. (1961) has been repeatedly underscored by replications of the classic study (Lewandowsky, 2011; Nosofsky, Palmeri, & McKinley, 1994). In all instances, models that did not incorporate some form of selective dimensional attention fit the data less well than models that did include selective attention (viz. ALCOVE; Kruschke, 1992). Moreover, the theoretically-expected allocation of attention across stimulus dimensions in the Shepard problems has recently been demonstrated more directly using eye-tracking analysis (Rehder & Hoffman, 2005a). There is now little dispute about the role of selective dimensional attention in category learning.

Dimensional attention can be readily linked to the executive-attention view of working memory: A crucial property of executive attention is the ability to suppress irrelevant information, preventing it from intruding into working memory (Engle & Kane, 2004; Kane et al., 2001, 2007). If dimensional attention operates analogously, it follows that it should come at a cost: Whereas learning about attended dimensions will be enhanced, unattended dimensions should be actively ignored, and the effects of that inattention should persist even when the task no longer demands it. This prediction has been confirmed by studies using dimensional relevance shifts (Kruschke, 1996b). In these tasks, an initially-learned category structure is replaced with a new stimulus-response mapping later in learning. The stimulus dimensions that are relevant to the second structure may or may not overlap with those relevant to the first structure. Whereas people can learn the new structure quite quickly if it involves the same diagnostic dimensions, people are quite slow to learn the second structure when previously irrelevant dimensions are suddenly made relevant to the task. More recently, Hoffman and Rehder (2010) have produced eye-tracking data confirming the attentional locus of the effect of relevance shifts on subsequent learning: People fixate less frequently on a previously irrelevant stimulus dimension even if it is suddenly made relevant to the altered category structure.

The apparent conceptual and empirical similarities between dimensional and executive attention suggest that WMC may be associated with attentional aspects of category learning. However, the limited available evidence has been negative. In a recent replication of the Shepard et al. (1961) study, Lewandowsky (2011) examined the dimensional attention-WMC link using structural equation modeling (SEM). The behavioral data were very clear: The classic ordering of the six problem types was reproduced, and at the level of individual participants, a model with a mechanism for selective attention (ALCOVE) fit the data better than a model without (the configural cue model; Gluck & Bower, 1988). SEM was then used to examine whether individual variation in performance could be captured by any of ALCOVE's parameters. A single latent variable was sufficient to account for the variation of parameters across all 6 problem types. Interestingly, the single parameter that loaded onto this latent variable (with estimates for each of the problem types constituting a separate manifest variable) reflected the rates at which direct stimulus-response associations were learned by the model. ALCOVE's attention learning parameter, by contrast, played virtually no role in explaining individual differences, despite the fact that the observed speed of learning across all problems was strongly related to WMC. This result runs counter to the expectation of a straightforward relationship between attention in category learning and WMC. Why did learning of dimensional attention fail to correlate with WMC?

We consider two possibilities in this paper. First, it could be the case that the Shepard problems are simply not sensitive enough to detect variation in people's ability to allocate dimensional attention: Only the Type I and II problems require deployment of selective dimensional attention; the remaining four problems require attention to all three stimulus dimensions. Given that people typically begin learning the Shepard problems in a diffuse attentional state (Rehder & Hoffman, 2005a), it follows that the dimensional attention learning demands across the suite of problems may have been minimal, thus obscuring any genuine relationship with WMC. On this account, a task that yields large and behaviorally unambiguous dimensional attention effects may fare better at revealing the link between dimensional attention and WMC. To this end, our first study used two association learning tasks that are widely thought to involve attentional learning; namely, blocking and highlighting (e.g., Kamin, 1968; Kruschke, 2009).[2] Second, it could be that the Shepard problems, and others commonly examined in the context of dimensional attention (e.g., the so-called 5-4 problem introduced by Medin & Schaffer, 1978, and discussed in detail by Rehder & Hoffman, 2005b, and Smith & Minda, 2000) do not sufficiently engage *executive* attention and thus WMC. Although dimensional attention seems to share many features with executive attention, there is an alternative account on which the two may actually be unrelated. Whereas dimensional attention operates over features of the stimulus, executive attention is usually construed as operating over abstract cognitive representations such as task goals (Engle & Kane, 2004; Kane et al., 2007). By implication, if controlling access to cognitive representations is central to executive attention as conceived by WM theoreticians, it is not dimensional, but representational attention in categorization that might be expected to relate closely to WMC. For example, Erickson (2008) aligned executive attention with the gating mechanism of the ATRIUM model of category learning (Erickson & Kruschke, 1998). In ATRIUM, representational attention is controlled by a gating mechanism, which determines whether a rule-based or exemplar-based category representation drives responding on a given trial. We discuss the implications of this view in detail when we introduce Experiment 2. For now, we suggest that executive attention may not contribute to the learning of a single distribution of dimensional attention, but may come into play when coordination of multiple subsets of knowledge, which map onto different category representations, is required. Because performance on the Shepard problems is readily explained in terms of distributions of dimensional attention over a single category representation, it follows that those problems may not have tapped executive attention. To examine this possibility, our second study used a knowledge restructuring task that required repeated changes in the way multiple category representations were coordinated during the task.

To foreshadow our principal results, we adduce support for the second alternative. That is, we fail to find any relationship between WMC and dimensional attention, as assessed by the magnitudes of blocking and highlighting effects. However, we uncover a significant relationship between WMC and the ability to change the way different category representations are coordinated during a task.

Attention and Associative Learning: Blocking and Highlighting

Since the discovery of associative blocking by Kamin (1968, 1969), there has been mounting evidence that attention plays a central role in associative learning and many contemporary models now include mechanisms for selective attention (Kalish, 2001; Kalish & Kruschke, 2000; Kruschke, 1996a, 2001b; Le Pelley, 2004; Mackintosh, 1975). In a typical blocking paradigm (summarized in Table 1), participants initially learn to associate some cue, A, with some outcome, X. Throughout this article, associative relationships are written using the notation, Cue → Outcome; hence A → X in Table 1. To ensure that A and X are distinguished from other cues and outcomes, F → Y trials are interleaved among the A → X trials during early learning. After these initial associations are learned, a new learning phase begins that involves two novel cue compounds presented with equal frequency, A.B → X, and C.D → Y. In the late training phase, A still predicts outcome X, as it did in the early learning phase; the only difference is that A has been paired with a *redundant* cue, B. Note also that cues B, C, and D have all been paired with their respective outcomes with equal frequency. If simple co-occurrence determines associative learning, it follows that the B → X association should be as strong as those between C → Y and D → Y. This prediction is assessed on final test trials that present the ambiguous compounds B.C $\rightarrow$? and B.D $\rightarrow$?, requiring participants to predict an unknown outcome. A very robust finding (e.g., Kruschke & Blair, 2000; Kruschke, Kappenman, & Hetrick, 2005; Le Pelley, Beesley, & Suret, 2007; Shanks, 1985) is that people strongly prefer outcome Y over outcome X; the implication is that the associations between C and Y and D and Y are stronger than the association between B and X.

Taken at face value, the blocking effect appears to show that very little has been learned about cue B. However, a dimensional attention account holds that people do learn something about cue B—they learn to *ignore* it. Although blocking per se does not necessitate an explanation based on attention—the principle of error-driven learning, as formulated in the classic model of Rescorla and Wagner (1972), suffices to produce blocking—the further consequences of blocking do.

To test the dimensional attention account of blocking, Kruschke and Blair (2000) elicited robust blocking effects using the procedure just described. In a subsequent third training phase, the blocked cue (B) was associated with a novel to-be-learned outcome (e.g., Z). If blocking were due to learned inattention, subsequent learning about a blocked cue should proceed at a slower rate compared to a control cue. By contrast, a purely error-driven account predicts equivalent learning rates due to the error signal introduced by the novel outcome. Kruschke and Blair's (2000) results were consistent with the dimensional attention account; the result has since been replicated several times (Kruschke, 2005b; Le Pelley et al., 2007). Further direct empirical support for the dimensional attention account of blocking was also adduced by Kruschke et al. (2005) who showed that gaze duration to blocked cues was reduced relative to non-blocked cues (see also Beesley & Le Pelley, 2011). These results clearly show that learned inattention is a major driving factor behind blocking, thus providing conceptual linkage with the executive attention said to underlie suppression of irrelevant information in working memory.

Another effect in associative learning, closely related to blocking, is attentional highlighting (Kruschke, 1996a, 2003, 2005b, 2009; Kruschke et al., 2005; see also, Gluck & Bower, 1988; Johansen, Fouquet, & Shanks, 2010; Kalish, 2001; Kalish & Kruschke, 2000; Medin & Edelson, 1988). Whereas in blocking, dimensional attention is directed *away* from a novel cue, highlighting involves directing dimensional attention *toward* a novel cue. In a typical highlighting design (also summarized in Table 1), participants are trained on a single association involving two cues and an outcome, I.pE $\rightarrow$ E, in an early training phase. In a later training phase, a new association is introduced, I.pL $\rightarrow$ L. Note that the associations are symmetrical in that each outcome is associated with a single perfect predictor, cues pE and pL, and shares a common imperfect predictor, I. If people learn the symmetry and exhibit statistically normative behavior, cues pE and pL should be equally predictive of outcomes E and L, respectively, whereas cue I should be regarded as non-predictive. In a subsequent test phase, people are presented with two critical stimuli, I $\rightarrow$?, and the ambiguous compound, pE.pL $\rightarrow$?. Interestingly, people show strong and conflicting choice outcome preferences for these stimuli. For I $\rightarrow$?, people exhibit a strong preference for outcome E. Conversely, for pE.pL $\rightarrow$?, people prefer outcome L. The effect illustrates a marked asymmetry in people's learning: Cue I is more strongly associated with outcome E, whereas the association between pL and L is stronger than the association between pE and E. A normatively irrelevant predictor (I) is deemed relevant, whereas one perfect predictor (pL) is apparently judged "more perfect" than another one (pE).[3]

Kruschke (2009) provided an attentionally-mediated order of learning account of highlighting, which postulates that highlighting occurs because *both* cues I and pE form associations with outcome E in the early learning phase. When the cue compound I.pL is introduced in later training, the pre-existing association between I and the erroneous outcome E introduces prediction error. The fastest way to eliminate this error is to shift dimensional attention away from cue I and onto pL *whenever the compound I.pL is presented*. The rapid attention shift serves to protect previous learning about the association between cue I and outcome E while accelerating learning about the relationship between pL and L.

The consequences of attentional highlighting can be assayed in the same way as for blocking. Kruschke (2005b) showed that subsequent learning about a previously highlighted cue was easier (i.e., more accurate) than a non-highlighted control cue (cf. Kruschke, 1996b; Kruschke & Blair, 2000). Kruschke et al. (2005) confirmed the attentional locus of the highlighting effect, showing increased gaze duration to highlighted cues relative to non-highlighted cues.

In sum, there is much evidence to suggest that both blocking and highlighting are bona fide attentional effects in associative learning. We examine the relationship between blocking, highlighting, and WMC in Experiment 1.

Experiment 1: Blocking & Highlighting

A relationship between WMC and blocking and highlighting effects would be consistent with the idea that WMC and blocking and/or highlighting engage a common form of attention. Failure to find any relationship would imply that blocking and highlighting do not rely on the same form of attention that is implicated in WMC.

Method

The experiment was spread over three 1-hour sessions scheduled at participants' convenience. Participants completed both the blocking and highlighting tasks in a single experimental session. The order of the blocking and highlighting tasks was determined by random allocation to one of 8 "modular sequences" described in detail later. The other two experimental sessions involved an unrelated categorization experiment. WMC was measured using the battery of 4 tasks presented by Lewandowsky, Oberauer, Yang, and Ecker (2010). WMC tasks were spread across the three sessions. All experiments were controlled by a Matlab program designed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Participants

A total of 140 people from The University of Western Australia community participated in the 3-session experiment either in exchange for course credit or remuneration at a rate of A$10 per hour. The sample was sufficiently large to allow structural equation modeling.

Modular Sequences

Following Lewandowsky (2011), we sought to strike a balance between maximizing experimental control via counterbalancing on the one hand, and reducing "method" variance on the other. We created 8 unique "modular sequences" to which participants were randomly assigned. Participants within a given modular sequence received the same stimulus sequence; that is, stimulus presentation order, and the mapping between stimulus features and abstract experimental design for the blocking and highlighting tasks were fixed across participants within a given modular sequence. Participants in different modular sequences encountered different training sequences and different stimulus feature mappings. The order in which participants completed the blocking and highlighting tasks was counterbalanced across modular sequences such that each task was first in four of the eight sequences.

Working Memory Capacity

Because working memory is a multifaceted construct (Oberauer et al., 2000, 2003), it is almost certain that a measurement based on a single task will overlook important aspects of the construct of interest. For example, Lewandowsky et al. (2010) analyzed a number of tasks intended to measure WMC and found variation in performance on these tasks to be a composite of variation due to a general WMC factor plus variation due to task-specific factors. To avoid the pitfalls of contaminating WMC measurement with task-specific variance, we used a heterogeneous WMC test battery in both experiments (Lewandowsky et al., 2010).

The four tasks in this battery addressed two content domains of WMC (verbal/numerical vs. spatial) in addition to two functional aspects of WMC (information storage in the context of online processing and integration of relational information). The tasks included operation span (OS; Turner & Engle, 1989), sentence span (SS; Daneman & Carpenter, 1980), spatial short term memory (SSTM; Oberauer, 1993), and memory updating (MU; Oberauer et al., 2000; Salthouse, Babcock, & Shaw, 1991).

Performance on the various working memory tasks have been found to load onto a single latent variable, thus providing a robust and reliable estimate of WMC. These tasks are described in detail in Lewandowsky et al. (2010) and we therefore do not restate those details here.

Blocking and Highlighting Tasks

The blocking and highlighting tasks were based on those used by Kruschke et al. (2005). In all cases, two copies of the abstract designs detailed in Table 1 were implemented in the experiment. For example, where A→X appears in Table 1, two versions, A1→X1 and A2→X2 were implemented in the actual experiment (see Tables 2 and 3 for details). Both tasks required determining the identities of "agents" sending a number of "coded messages". The coded messages were constructed from a set of stimulus words that were randomly sampled from the same pool of 20 candidate words used by Kruschke et al. (2005). Candidate words were 5-letter nouns with familiarity, imagability, and concreteness ratings above 500, as recorded in the MRC Psycholinguistics Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). Each modular sequence involved a different random sample of stimulus words. Within each modular sequence, there was no overlap between the set of stimulus words used for the blocking task, and the stimulus set for the highlighting task. The mapping between implemented and abstract response alternatives (i.e., the "senders" of the coded messages) was invariant across subjects, but differed between blocking and highlighting tasks.

Procedure

Participants completed all tasks in a sound-attenuated testing booth. The basic trial structure for the blocking and highlighting tasks was very similar. An example blocking trial is presented schematically in Figure 2. Participants indicated their response via mouse click in one of the response boxes below the fixation point. During training, corrective feedback was provided directly following a response in the location previously occupied by the response prompt: 'CORRECT! The message was sent by . . . ' or 'WRONG! The message was sent by . . . '. For test stimuli, which were never paired with any outcome, the message 'Response recorded' was presented instead. Feedback remained visible until participants terminated the trial by clicking a button labeled "Next" that appeared in the center of the response alternatives. Appearance of the feedback and the "Next" button was separated by a mandatory 1 second study period. A 500 ms blank interval separated trials.

The training regime for the blocking task involved an early phase followed by a late phase. Each phase involved 80 training trials, divided into 10 8-trial blocks. Each training block comprised a random permutation of the relevant subset of stimuli detailed in Table 2. The training phases were followed by a test period that involved all of the stimuli included during the late training phase (which were presented twice each) in addition to a set of previously unencountered test stimuli (presented once each). Test stimuli were presented in a random order determined by the particular modular sequence. Performance on the training stimuli presented during the test phase was used to assess extent of learning. The test stimuli were divided into two classes. B.(C/D) stimuli paired the blocked cue B with one of the control cues, C or D. Responses to these stimuli indicate whether the C→Y and D→Y associations are stronger than the B→X associations, and thus test directly for blocking. By contrast, A.(C/D) stimuli pair the blocking cue A with the control cues. Responses to these items reflect the strength of the A→X association relative to the C→Y and D→Y associations.

The training regime for the highlighting task followed the canonical design proposed by Kruschke (2009), which equates base rates for all training stimuli, and involved three training phases (see Table 3 for details). In total there were 168 training trials (excluding the training stimuli interleaved among test stimuli, which were used to assess learning of the training set). As with the blocking task, the order of trials was randomly permuted within each block. Test stimuli were ordered by randomly permuting the entire stimulus set. The two classes of test stimuli were designed to assess the association strengths between the I, pE, and pL cues and their trained outcomes. I._ test stimuli involved I cues presented in isolation to determine whether they were differentially associated with outcome E. The pE.pL test stimuli paired the two perfect predictors of outcomes E and L to determine whether the pL→L association dominated the pE→E association by virtue of attentional highlighting of the pL cue.

Results

Data Screening

To be included in the final analysis, a participant needed to have (1) completed all WMC tests, (2) performed better than 70% on the processing components of the two complex span tasks (OS and SS), and (3) performed significantly better than chance on the training stimuli that were interleaved with test items in both the blocking and highlighting tasks. For the blocking task, the chance cutoff was 8 correct responses out of the 16 relevant trials (given $p = .25$ and binomial assumptions). For the highlighting task, correct responses on at least 6 of the 8 training trials were needed to exceed chance. Only three participants failed to meet the performance criteria on either the blocking or highlighting tasks. An additional 16 participants were removed from the analysis because of either failing to complete all test sessions (9 participants), or performing worse than 70% on at least one of the processing components in the span tasks. In total, data from 121 participants remained for analysis.

Working Memory Battery

WM performance was scored using a partial-credit scheme (cf. Conway et al., 2005). For instance, a participant 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 4; they are consonant with those obtained during previous applications of the battery (Craig & Lewandowsky, in press; Ecker, Lewandowsky, Oberauer, & Chee, 2010; Lewandowsky, 2011; Lewandowsky et al., 2010; Lewandowsky, Yang, Newell, & Kalish, under review).

Blocking and Highlighting Effects

Our analysis of the blocking and highlighting data follows that of Kruschke et al. (2005). Overall accuracy on the 16 training stimuli presented at test was at ceiling, $M = .97$, indicating robust learning of the training set. There was no obvious difference in learning between the A.B and C.D training stimuli, both Ms$= .97$.

We next examined choice preferences for the two classes of test stimuli. For the B.(C/D) class stimuli, which paired the blocked cue B with control cues C and D, participants chose the response option associated with the control cue (Response Y) 59% of the time. The response associated with the blocked cue (Response X) was chosen only 35% of the time. The remaining 6% of responses were distributed across the other two response options. We counted the number of times each participant gave the relevant Y response minus the number of times that participant gave the relevant X response, divided by the number of B.(C/D) trials (i.e., 16). This yielded an index of choice preference for the B.(C/D) class of stimuli, which we denote *DmBc*, for D minus B for choice. The mean *DmBc* value was 24%, which was significantly greater than 0, $t(120) = 7.24, p < .001, r^2 = .30$, indicating a robust blocking effect.

To confirm that the effect could not be attributed to a bias to choose the control response when faced with conflicting cues, we examined choice preferences for Class A.(C/D) test items, which paired cue A with the two control cues. Responding reflected a differential preference for the outcome associated with cue A (Response X), which was selected 68% of the time. The response associated with the relevant control cue (Response Y) was selected on only 30% of trials. The remaining 2% of responses were distributed across the other two response options. We computed analogous component measures of choice preference by counting, for each participant, the number of X responses minus the number of Y responses, divided by the number of A.(C/D) trials (i.e., 16). This choice preference index for A.(C/D) stimuli is denoted *AmCc*, for A minus C for choice. The mean *AmCc* was 37%, which was significantly greater than 0, $t(120) = 11.42, p < .001, r^2 = .52$, indicating that the blocking effect was not due to a general preference for the control outcome.

We analyzed the highlighting data in a similar manner. To confirm that the training set was learned, we examined response accuracy to the I.pE and I.pL training stimuli presented during the test phase. Overall accuracy on these 8 training stimuli was at ceiling, $M = .95$. Comparably high levels of learning were achieved for the I.pE and I.pL training stimuli, Ms = .97 and .94, respectively.

We next examined choice preferences for the two classes of test stimuli. For the pE.pL test stimuli, which paired the perfect predictors for outcomes E and L, participants chose the response option associated with the pL cue (Response L) 64% of the time. The response associated with the early trained cue pE (Response E) was chosen only 33% of the time. The remaining 3% of responses were distributed across the other two response options. As with the blocking results, we counted, for each participant, the number of relevant L responses minus the number of relevant E responses, divided by the total number of pE.pL trials (i.e., 8). We refer to this choice index as *pLmpEc*, for pL minus pE for choice. The mean *pLmpEc* was 31%, which was significantly greater than 0, $t(120) = 7.46, p < .001, r^2 = .32$, reflecting a robust highlighting effect.

To determine whether the imperfect cue I became differentially associated with outcome E, we examined responses to the I._ class of test stimuli. Outcome E was chosen 66% of the time, compared to outcome L, which was chosen only 30% of the time. The remaining 4% of responses were distributed across the remaining two response options. A choice index, *Ic*, was computed by counting, for each participant, the number of E responses minus the number of L responses, divided by the total number of I._ trials (i.e., 8). The mean *Ic* value was 36%, which was significantly greater than 0, $t(120) = 9.09, p < .001, r^2 = .41$, indicating that cue I was differentially associated with outcome E.

Covariation of Effects

Following Kruschke et al. (2005), we obtained general indices of the overall blocking and highlighting effects by summing the subcomponent choice measures. Thus, the blocking effect was indexed by $DmBc{+}AmCc$, and highlighting was indexed by $pLmpEc{+}Ic$. Before investigating the patterns of correlations between blocking and highlighting, in addition to the correlations among their component measures, we corrected the empirical correlations for attenuation due to measurement error (Spearman, 1904). The correction for attenuation compensates for the fact that empirical correlations are systematically biased downward when data are collected in the presence of measurement error, and can result in corrected correlations greater than unity (see Charles, 2005, for discussion). The corrected correlation is obtained by dividing the empirical correlation by the geometric mean of the reliabilities of the variables being correlated,

$$r_{xyc} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}. \tag{1}$$

The need to correct for attenuation is greater with lower test reliabilities (Fan, 2003). We computed Cronbach's αs both for the blocking and highlighting tasks as a whole, and separately for the subclasses of test stimuli. For example, in the blocking task, the response on each trial was coded as either consistent with blocking (+1, observed response was associated with the control cue), inconsistent with blocking (-1, observed response was associated with the blocked cue), or neither (0, observed response was not associated with either cue presented on that trial). Table 5 summarizes the reliabilities used to correct the raw empirical correlations.

We report both raw (written as r_{emp}) and corrected correlations (written as r_{cor}) below, p values are included for the raw correlations. For the corrected correlations, we report 95% confidence intervals based on methods described by Charles (2005). We first examined the correlation between the two component measures of blocking, which were positively correlated, $r_{\text{emp}} = .49, p < .001$. Correcting for attenuation yielded, $r_{\text{cor}} = .73, 95\%\ CI = [.43, 1.25]$. For the component highlighting measures, the results were less clear-cut. The raw empirical correlation was marginally significant, $r_{\text{emp}} = .16, p < .08$, and although correcting for attenuation increased the magnitude of the correlation, the 95% confidence interval narrowly included 0, $r_{\text{cor}} = .31, 95\%\ CI = [-.04, .73]$. With the exception of the fact that the correlation between the component highlighting measures was only marginal, we replicated the general pattern of correlations reported by Kruschke et al. (2005).

We next examined whether the magnitudes of the overall blocking and highlighting effects covaried within individuals. Curiously, we found no evidence of any correlation between the magnitude of an individual's blocking and highlighting effects, $r_{\text{emp}} = .04, p = .66$, even after correcting for attenuation, $r_{\text{cor}} = .06, 95\%\ CI = [-.22, .36]$. This diverges from the results of Kruschke et al. (2005), who found a significant positive correlation of .38. To confirm that our failure to find any correlation between blocking and highlighting was not due to "impurities" introduced by the *AmCc* and *Ic* measures, we correlated the component performance indices that map most directly onto the individual blocking and highlighting effects (viz. by correlating *DmBc* with *pLmpEc*). This correlation was also nonsignificant, $r_{\text{emp}} = .10, p = .29$, even after correcting for attenuation, $r_{\text{cor}} = .17, 95\%\ CI = [-.15, .52]$. Given that our sample size was roughly three times larger than that used by Kruschke et al. (2005), we are reluctant to ascribe the absence of a correlation to a lack of power in our study.

We investigated the lack of correlation between blocking and highlighting effects via computational modeling with EXIT (Kruschke, 2001b). We summarize the details of the modeling here; full details are reported in Appendix A. Our hypothesis was that the absence of a blocking-highlighting correlation was due to differential involvement of dimensional attention across the two tasks; namely, that dimensional attention was necessary to produce highlighting, but not blocking. We adopted a nested model approach, whereby we fit EXIT to the combined data from the blocking and highlighting tasks. That is, the model was forced to fit both data sets with a common set of parameter values. We then relaxed this constraint and allowed the model's attention parameter values to differ between the tasks. If the more flexible model produces a significant improvement in fit to data, accounting also for the increased complexity of the more flexible model, it is concluded that the additional flexibility is warranted by the data.

We found that EXIT could only account for the data quantitatively when its attention parameters differed across the blocking and highlighting tasks. Inspection of the parameter values revealed that the model utilized both attentional and associative mechanisms for fitting the highlighting data. For fitting the blocking data, the model effectively shut off its attention shifting and attention learning mechanisms, and instead relied only on associative learning (see Appendix A for full details). The modeling is consistent with previous theoretical claims that dimensional attention need not be invoked to explain blocking per se (Rescorla & Wagner, 1972). It also confirms our hypothesis that the absence of a blocking-highlighting correlation was due to differences in the attentional demands of the tasks.

Although our conclusions about performance in the blocking task run counter to those of Kruschke et al. (2005), it is well known that dimensional attention is not necessarily required to explain blocking per se. Our results merely suggest that identification of the determinants of whether a given example of blocking arises from associative or attentional factors awaits future research. These issues do not impinge on our primary focus of how individual differences in WMC relate to variation in blocking and highlighting performance. We now report structural equation modeling that addresses the key individual differences issues of interest.

Structural Equation Modeling

Measurement Models. The measurement model for WMC included a single latent variable, which we label *WMC*. Consistent with previous applications of the WMC battery (Lewandowsky et al., 2010), the fit of the model was improved by freely estimating the correlation between the error terms associated with the OS and SS tasks. The benefit of estimating the correlation between these error terms derives from the fact that the two tasks are structurally quite similar. OS examines memory in the context of a numerical processing task, whereas SS examines memory in the context of a verbal processing task. The resulting model provided an excellent fit, $\chi^2(2) = 0.0$; $CFI = 1.0$; $RMSEA = 0.0$, 90% CI = [0.0, 0.0], $SRMR = .0054$. The loadings of the four manifest variables are shown in the bottom row of Table 4.[4]

We explored several measurement models for the blocking and highlighting tasks. The first model involved 6 manifest variables, four of which tracked learning performance; namely, the (log transformed) total number of errors in the early and late phases of the blocking and highlighting tasks. (For the highlighting task, we pooled the two late phases into one; see Tables 2 and 3). The remaining two manifest variables tracked the attentional aspects of the tasks; namely the overall measures of blocking and highlighting. Means and variances of the manifest variables are reported in Table 6.

Because of the lack of correlation between the blocking and highlighting measures, we were unable to investigate a two-factor model that considered learning separately from dimensional attention. Instead, we investigated the loadings of each manifest variable onto a single latent variable. All four learning indicators loaded significantly onto the latent variable, in contrast to the attentional manifest variables (i.e., blocking and highlighting), both of which failed to load.[5] Accordingly, we label the latent variable in this model *Errors*, and interpret the single-factor in terms of learning, not attention. The fit of the single-factor model was very good, $\chi^2(9) = 7.7$; $CFI = 1.0$; $RMSEA = 0.0$, 90% CI = [0.0, 0.092], $SRMR = .0455$.

Structural Model. The two measurement models were combined into a final structural model, shown in Figure 3. (Complete correlation matrices for both experiments are provided in Appendix C). The structural model shows how learning accuracy in both tasks (captured by the latent variable, *Errors*) relates to the *WMC* latent variable measured by the WMC battery. The model fit very well with the weights involving blocking and highlighting set to 0, $\chi^2(35) = 30.3$; $CFI = 1.0$; $RMSEA = 0.0$, 90% CI = [0.0, 0.053], $SRMR = .0554$.

The significant negative correlation, r =-.44, between the latent variables *WMC* and *Errors* indicates that higher WMC was associated with *fewer* errors during learning in both the blocking and highlighting tasks. However, the lack of a measurement model for a common attention construct prevents us from drawing conclusions about the relationship between a putative dimensional attention construct and WMC. To address this issue, we used the WMC measurement model to examine the relationship between the WMC latent variable and each attention measure in isolation. For this analysis, tantamount to investigating regression coefficients relating the behavioral indices of dimensional attention (i.e., measures of the highlighting effect) to the *WMC* latent variable, we took the measurement model for WMC, fixed all manifest variable loadings, then added the highlighting index as well as its component indices, Ic and $pLmpEc$, as additional indicators of the *WMC* latent variable. None of the highlighting measures loaded significantly onto the *WMC* latent variable (βs for highlighting, Ic, and $pLmpEc$ = -0.06, -0.07, -0.02, respectively, ps $> .43$). Thus, we conclude that there is no evidence in our data of a relationship between dimensional attention and WMC.

An apparent complication in interpreting the SEM results from Experiment 1 arises from the fact that Lewandowsky (2011) found association learning in ALCOVE to correlate with WMC. That is, in his study, the association-learning parameter in ALCOVE increased with people's WMC when the model was fit to the data from individual subjects. Given that our blocking effect was arguably associative in nature—as revealed by modeling with EXIT—one might have expected blocking, but not highlighting, to load onto the latent variable in our structural equation model. Upon further investigation, this lack of loading turned out to reflect the fact that the conventional measure of blocking did not adequately reflect the associative history between cues and outcomes. We report the results of this extended analysis in Appendix B, noting that once associative history is accounted for, blocking performance loads onto the *Errors* latent variable in the expected manner.

To summarize the SEM results, our findings are completely consonant with related precedent (Craig & Lewandowsky, in press; Lewandowsky, 2011; Lewandowsky et al., under review). All manifestations of association learning are related to WMC whereas there is no obvious link between dimensional attention and WMC.

Implications

The structural equation modeling converges on a clear conclusion: Dimensional attention and WMC are distinct theoretical constructs. Individual differences in learning of both tasks in Experiment 1 was characterized by a single latent variable that was significantly correlated with WMC. By contrast, we failed to find any relationship between blocking and highlighting, the latter of which was confirmed to be mediated by dimensional attention (see Appendix A), and WMC. Both aspects of our results mesh well with the results of Lewandowsky (2011) who found minimal attentional involvement in the modeling of individual differences in learning of the Shepard problems.

The results of Experiment 1 are compatible with two competing conclusions: One possibility is that executive attention as defined in the working memory arena is entirely distinct from attention in category learning (viz. dimensional attention). Another possibility is that executive attention is involved in category learning, but is related instead to representational attention, which is implicated in more complex categorization tasks involving the coordination of multiple category representations. We investigate this possibility in Experiment 2.

Experiment 2: Knowledge Restructuring

Whereas Experiment 1 investigated the link between WMC and dimensional attention, Experiment 2 focuses on a possible relationship between WMC and representational attention (Erickson, 2008; Erickson & Kruschke, 1998, 2002b). In categorization, representational attention is invoked in situations where different subsets of stimuli are classified on the basis of different representational formats; for example, some stimuli may be classified according to a rule, whereas an exemplar-based representation might handle a different subset of stimuli.

The distinction between dimensional attention on the one hand and representational attention on the other is illustrated schematically in Figure 4. Each panel depicts a scheme for relating stimulus inputs to category outputs. In the top panel, dimensional attention affects the pattern of exemplar activation by selectively enhancing inputs along dimensions $D1$ and $D2$—increased dimensional attention weights are drawn in boldface in the figure. Category responses are determined by the stimulus-response mapping, which in the top panel is a matrix of exemplar-to-category associations. The bottom panel illustrates a model with both dimensional and representational attention. The effects of dimensional attention are the same as those in the top panel. Representational attention gates the associations between individual exemplar nodes and the various representational formats available to the model: Exemplars 1 and 2 gate access to representation $R1$, whereas Exemplars 3 and 4 gate access to representation $R2$. Category responses are determined by the output of the selected representation.

In the figure, the exemplar-to-category association matrix in the dimensional-attention model (top panel) is assumed to be identical to $R1$ in the representational attention model (bottom panel). Models with only dimensional attention can be viewed as special cases of representational-attention models: For example, the model shown in the bottom panel can be reduced to the model in the top panel if the distribution of representational attention is such that each exemplar gates access to a common representation (e.g., if all exemplar nodes selectively accessed $R1$). Thus, models with representational attention can be seen as supersets of models that only comprise dimensional attention.

One category learning model that implements representational attention is ATRIUM, which minimally learns to associate stimuli with either a rule or exemplar representation, one or the other of which determines classification (Erickson & Kruschke, 1998). The pattern of activation across ATRIUM's exemplar nodes is determined by the model's allocation of dimensional attention. Exemplar activations determine, via a representational-attention "gate," whether a rule representation or exemplar-based representation is accessed and subsequently used to generate a category response. Representational attention is therefore inextricably related to the *coordination* of multiple category representations, which individually constitute subsets of partial category knowledge. Within ATRIUM, dimensional and representational attention are distinct, but related constructs. They are distinct because dimensional attention is involved in similarity computations and the weighting of stimulus input, whereas representational attention is involved in the selection of a category representation. They are related in that shifts of representational attention can be induced by lower-level shifts of dimensional attention (Sewell & Lewandowsky, 2011).

The notion of representational attention, and the attendant coordination of partial knowledge, is readily explored in so-called knowledge restructuring tasks, which involve switching between a number of candidate response strategies (e.g., Kalish, Lewandowsky, & Davies, 2005; Lewandowsky, Kalish, & Griffiths, 2000; Little, Lewandowsky, & Heit, 2006; Sewell & Lewandowsky, 2011). For example, the study by Sewell and Lewandowsky (2011) involved the coordinated application of two partial categorization rules to different subsets of stimuli. Their category structure, which was also used in our Experiment 2, is shown in Figure 5 along with two example stimuli.

The stimuli were rectangles that varied along two dimensions: The height of the rectangle, and the position of a vertically oriented bar along the base of the stimulus. The rectangle width was fixed. Thus, each point in Figure 5 corresponds to a different configuration of rectangle height and bar position. The example stimulus shown on the left side of the figure was sampled from the bottom left region of the category space (short rectangle with bar position on left), whereas the example stimulus on the right side of the figure was sampled from the top right region of the space (tall rectangle with bar position on right).

Experiment 2 was broken down into different phases comprising training and test blocks. On each trial during training blocks, participants were required to categorize a single stimulus from among the set of training stimuli (represented as filled diamonds in the figure); responses were immediately followed by corrective feedback. On each trial during test blocks, participants categorized a single stimulus drawn from the set of transfer stimuli (represented as open squares in the figure); no feedback was provided during any test block.

Several features of the category structure make it diagnostic of the use of representational attention. Note that there are two separate clusters of training stimuli, one in the bottom left corner of the category space, the other in the top right. Each cluster comprises 20 Category A stimuli and 20 Category B stimuli, which are separated from one another by a partial category boundary (along one of the stimulus dimensions). With respect to the left training cluster, Category A stimuli are situated below the partial boundary, whereas Category B stimuli are situated above the boundary. With respect to the right training cluster, Category A stimuli are placed to the left of the boundary, whereas Category B stimuli are placed to the right of the boundary. Note that the two partial boundaries cannot be integrated in a coherent manner—neither partial boundary can be extended in a way that permits accurate classification of the training cluster straddling the other partial boundary. Thus, the category structure is ideally suited to observe the coordination and selective application of multiple partial rules.

In addition to the two stimulus dimensions depicted along the x and y axes in Figure 5, a third binary context dimension (instantiated by stimulus color) was systematically mapped onto the two training clusters. For example, all training stimuli in the cluster straddling the left partial boundary may have been presented in red, whereas stimuli from the right training cluster may have been presented in green. The inclusion of this third context dimension enables at least two ways in which the partial categorization rules could be coordinated and applied to the broader set of transfer stimuli. Under a knowledge partitioning (KP) strategy, context is used to determine whether the left or right partial boundary is applicable. By contrast, under a context insensitive (CI) strategy, the position of the stimulus along the x axis (i.e., whether a stimulus is on the left- or right-hand side of the space) determines which rule to apply.[6] Both strategies can support perfect performance during training (stimuli represented by filled diamonds in the figure), but they lead to qualitatively different patterns of performance on the transfer test (open squares), when the entire set of transfer stimuli are presented once in each context. Figure 6 illustrates idealized versions of these two response profiles.

Sewell and Lewandowsky (2011) instructed participants to use either the KP or CI strategy at the outset. In each case, people were told that for each stimulus, only a single dimension was required for categorization, but that the relevant dimension depended on the stimulus itself. Thus, participants were told that bar position determined category membership for some stimuli, but rectangle height determined category membership for others. People instructed to use the KP strategy were told that context (i.e., stimulus color) reliably indicated whether height or bar position was diagnostic. By contrast, people instructed to use the CI strategy were told that bar position (i.e., whether the bar was on the left- or right-hand side of the stimulus) indicated which dimension was diagnostic. Thus, people were told to initially use one stimulus dimension to gate subsequent use of another stimulus dimension to perform categorization. Participants completed an initial training phase, followed by a transfer test, performance on which was diagnostic of strategy use. Participants were then instructed to use the contrasting strategy—which had never been mentioned before—before completing another transfer test. People who were initially told to gate rule use on the basis of context were informed that bar position should be used to determine the relevant stimulus dimension, and vice-versa.

In a final testing phase, instructions were yet again reversed and people reverted to their original strategy. Thus, participants performed the task either under KP–CI–KP instructions or in a CI–KP–CI order. Several aspects of the results are noteworthy. People were able to rapidly and repeatedly recoordinate their partial knowledge in response to instructions: Whenever they received the KP instructions, their transfer profiles resembled those in the bottom panels of Figure 6, and whenever they received CI instructions, their profiles resembled the top panels. Remarkably, people could shift between strategies without requiring any training on the novel strategy—thus, learning under CI instructions enabled people to switch to the KP strategy, and vice versa, simply in response to a written hint and without feedback-driven learning. We used the same instructional regime in Experiment 2.

Sewell and Lewandowsky (2011) showed that this repeated, fluid, and near-instantaneous knowledge restructuring was incompatible with a slow associative or attentional learning process, as instantiated in ALCOVE (Kruschke, 1992). Instead, the results were quantitatively modeled by a version of ATRIUM that was equipped with multiple rule representations (cf. Yang & Lewandowsky, 2004). In ATRIUM, knowledge restructuring was modeled by a shift in dimensional attention, which in turn elicited a shift in representational attention. Specifically, when the context dimension had a high dimensional attention weight, gating of the rule representations was based on context, thereby implementing the KP strategy; when the x dimension had a high dimensional attention weight, rule selection was determined by the stimulus position along the x axis, irrespective of context, thus implementing the CI strategy. Sewell and Lewandowsky (2011) further showed that the content of the rules underpinning the KP and CI strategies was virtually identical: Associations within the rule modules were highly correlated regardless of which response strategy was used at transfer. The fact that common rule-based knowledge was used to instantiate both strategies provides converging evidence that the observed restructuring was driven solely by a shift in representational attention.

The controlled and volitional aspect of the knowledge restructuring reported by Sewell and Lewandowsky (2011) distinguishes it from Experiment 1 and the category learning tasks investigated by Lewandowsky and colleagues (Craig & Lewandowsky, in press; Lewandowsky, 2011; Lewandowsky et al., under review): Switching strategies did not involve a rapid shift of dimensional attention provoked by the stimulus (as in highlighting), nor was there a performance-related error signal to direct gradual learning of dimensional attention (as in blocking and typical supervised categorization). Whereas the attentional factors involved in blocking (when it involves attention), highlighting, and supervised categorization tasks can be viewed as "bottom-up," knowledge restructuring and recoordination can occur in the absence of changes in the stimulus, and are thus better characterized as involving "top-down" attentional control.

We suggest that control over the coordination of multiple rule representations relates closely to notions of executive attention as discussed in the working memory literature. Application of one cognitive strategy over another that affords equal performance requires a high degree of top-down control involving a combination of selectively maintaining one set of task goals associated with one strategy while selectively suppressing a competing set of goals associated with another strategy. Theoretically, this maps well onto contemporary notions of executive attention (e.g., Engle, 2002; Engle & Kane, 2004; Engle et al., 1999; Kane et al., 2001, 2007). We thus expected that the extent to which people can restructure their knowledge and switch categorization strategies should correlate with measures of WMC.

Method

Experiment 2 was similar to the first study, with the primary difference being the nature of the categorization task. In the first categorization session, we initially instructed participants to learn either the KP or CI strategy by telling them that either the context or bar position indicated whether the stimulus could be categorized on the basis of rectangle height or bar position. The category structure was identical regardless of instruction condition.

People then completed a training phase, which was followed by a transfer test. Afterwards, participants were instructed to switch response strategies, and had to categorize the entire set of transfer stimuli without having practiced the new strategy. That is, people who were initially told to select rules on the basis of context, were now told to select rules on the basis of bar position, and vice-versa.

In the second categorization session, participants had an opportunity to practice the new strategy, followed by another transfer test. Upon completion of the transfer test, participants were again suddenly instructed to switch response strategies before completing a fourth and final transfer test (i.e., in the second test following Categorization Session 2 they were instructed to reinstate their original response strategy from the first session).

Participants and WMC measurement

A total of 106 people from The University of Western Australia community participated either in exchange for course credit or remuneration at a rate of A$10 per hour. Working memory capacity was assessed in the same way as in Experiment 1 in the first of three experimental sessions.[7]

Participants were randomly allocated to one of 8 modular sequences that were analogous to those from Experiment 1. Each sequence involved a unique pre-loaded trial order and stimulus assignment. The mapping between color and context was counterbalanced across modular sequences. Half of the modular sequences required participants to learn the KP strategy first.

Procedure

The categorization part of the study involved two sessions carried out on separate days. Each session involved a training period followed by two successive transfer tests. For clarity, we refer to the transfer tests from Categorization Session 1 as Tests 1 and 2, and those from Categorization Session 2 as Tests 3 and 4. The task involved categorization of rectangle stimuli that varied along two dimensions: The height of the rectangle (y dimension in Figure 5) and the horizontal offset of a bar located along the base of the rectangle (x dimension in Figure 5). Depending on which modular sequence a participant was assigned to, they were instructed to learn either the KP (KP-first condition) or CI

(CI-first condition) strategy; 53 people were assigned to each condition. For the KP-first condition, participants were informed that context determined whether the x or y dimension was relevant for a given stimulus. For the CI-first condition, participants were told that the position of the bar offset (on either the left- or right-hand side of the stimulus) determined whether the x or y dimension was relevant for categorizing a particular stimulus. No information about the positioning of the partial rule boundaries along each dimension was given (i.e., there was no mention of specific "cutoff" values).

Training involved 6 40-trial blocks; each training stimulus was presented once per block with presentation order determined by the random permutation for the particular modular sequence. For Categorization Session 1 only, we allowed for an early exit from training if a participant made 40 consecutive correct responses. The earliest possible exit was after completion of 4 complete training blocks (160 trials). Because this is a fairly stringent criterion, it seemed unlikely to pose any problems in interpreting subsequent test performance (Tharp & Pickering, 2009). For the purposes of assessing training accuracy, it was assumed that participants who met the early exit criterion achieved perfect performance for the remainder of the training period. There was no early exit from Categorization Session 2 training, which was only 2 blocks in duration (80 trials).

Results

Data Screening

We applied the same retention thresholds as in the first study. Three participants (two from the CI-first condition, one from the KP-first condition) failed to perform better than chance in the categorization task (i.e., assuming a binomial response model, greater than 65% in the final training block of the first session). A further two participants scored less than 70% on at least one of the WMC processing tasks. A further participant was removed due to incomplete WMC data, yielding a final sample of 100 participants for analysis.

Working Memory Battery

Table 7 shows summary statistics for the WMC battery. Not surprisingly, the data mirror the results of the first experiment.

Categorization Analysis

Training Performance. Training performance in Categorization Session 1 was very similar in both the KP-first and CI-first conditions. Performance was highly accurate overall, with 32 people from the CI-first condition and 32 people from the KP-first condition achieving the early exit criterion. For the remaining participants in each condition, performance on the final Categorization Session 1 training block was highly accurate, $M = 89\%$ for both the CI-first and KP-first conditions. Training performance in the final Categorization Session 2 training block was comparably high (Ms $= 95\%$ and 96% for the CI-first and KP-first conditions, respectively). Thus, we conclude that the training sets were learned extremely well—and to an equal extent—in both conditions and under both response strategies.

Strategy Differences in Test Performance. To confirm the effectiveness of the initial instructions, we compared performance of the KP-first and CI-first conditions on Transfer Test 1. The averaged response profiles for each condition are presented in Figure 7. Each square in the figure shows the probability of generating an "A" response, $P(A)$, for that stimulus. It is clear that the CI-first condition relied on the CI strategy, whereas the KP-first condition used a context-sensitive knowledge partitioning strategy.

To compare performance of the two conditions statistically, we first divided the space into four diagnostic regions by aggregating across transfer stimuli in the four quadrants of the space (i.e., the regions defined by the dashed lines and labeled numerically in Figure 5) for each context. We then averaged the aggregated response data across participants. A 2 (Condition) $\times$ 2 (Context) $\times$ 4 (Quadrant) between-within ANOVA returned a significant 3-way interaction, F (3, 294) = 92.89, MS_e = .01, $p < .001$, $\eta_p^2 = .29$, reflecting the differential sensitivity to context between conditions. Follow-up Condition $\times$ Quadrant between-within ANOVAs on performance within each context revealed condition-specific patterns of responding across Quadrants; significant interactions were observed in the left context, $F(3, 294) = 47.03$, $MS_e = .02$, $p < .001$, $\eta_p^2 = .32$, and the right context, $F(3, 294) = 89.05$, $MS_e = .03$, $p < .001$, $\eta_p^2 = .48$. The pattern of results demonstrates that the hints were effective at determining the manifest categorization strategy in the first transfer test.

Performance on all other transfer tests corresponded very closely to the two patterns shown in Figure 7. To efficiently capture changes in strategy use across all four transfer tests, we introduce a context-sensitivity measure that tracks usage of the KP and CI strategies. This measure was computed by taking the average item-wise difference in $P(A)$ for stimuli that would be categorized differently between contexts if the KP strategy were perfectly applied (see Figure 6). Responses for the four stimuli in the bottom right corner of the space were reverse coded for this analysis, as categorization of these stimuli changes in the opposite manner to all other relevant stimuli as a function of context under the KP strategy. Thus, context sensitivity ranges from -1 to $+1$, with positive values reflecting response patterns consistent with application of the KP strategy, and values near 0 reflecting usage of the CI strategy. Accordingly, Test 1 performance in the KP-first condition was associated with a high context sensitivity score, $M = .87$, whereas the CI-first condition was associated with a low score, $M = .10$. Figure 8 plots context sensitivity as a function of transfer test for both conditions. It is clear that the hints had immediate and opposite effects on performance across the two conditions.

Also of note is the apparent asymmetry of knowledge partitioning between the two groups. Participants in the CI-first condition, who were initially trained to ignore context, appeared to exhibit weaker knowledge partitioning than participants in the KP-first condition — compare the context sensitivity scores of the CI-first condition in Tests 2 and 3 to those of the KP-first group in Tests 1 and 4. By contrast, there was no discernible difference between the groups when implementing the CI strategy. This asymmetry implies a selective difficulty with shifting from the CI to the KP strategy that is absent in the reverse direction. We interpret this difference in terms of dimensional relevance shifts: It is known that it is easier to attend to a previously-relevant stimulus dimension than it is to attend to a previously-ignored dimension (Hoffman & Rehder, 2010; Kruschke, 1996b). We first note that under both strategies, the rectangle height and bar position dimensions were necessary for categorization. The difference between conditions was whether the context dimension was to be used *in addition* to the other stimulus dimensions. For people in the KP-first condition, all stimulus dimensions were relevant during initial training, and so restructuring to the CI strategy involved using a previously relevant stimulus dimension (bar position) for a new purpose (rule selection). By contrast, people in the CI-first condition were effectively instructed to ignore the context dimension during initial training (i.e., context was not involved in either rule selection or categorization). Thus, restructuring to the KP strategy involved having to use a previously irrelevant stimulus dimension for the purposes of rule selection.

Sewell and Lewandowsky's (2011) modeling of a similar pattern of results with ATRIUM provides convergent support for the relevance shift interpretation of the asymmetrical knowledge restructuring in Figure 8. Sewell and Lewandowsky modeled knowledge restructuring by implementing changes in ATRIUM's distribution of dimensional attention, which in turn caused shifts of representational attention. There was an asymmetry in changes in the attentional loading on the context dimension that mirrored the context sensitivity data. For example, between Tests 1 and 2, the attention weight on context dropped from .74 to 0 for the KP-first condition, but increased only from .04 to .65 for the CI-first condition. Thus, the ability of the CI-first condition to attend to the context dimension appeared weaker than the ability of the KP-first condition to ignore it in Test 2. It is perhaps puzzling that participants' additional training between Tests 1 and 2 did not further increase context sensitivity in the CI-first condition. Given the difficulty in attending to a previously irrelevant stimulus dimension, we suggest that the absence of an error signal during training may be the reason the CI-first condition was unable to achieve the same level of context sensitivity as the KP-first group when implementing the KP strategy—both the KP and CI strategies would result in perfect training performance.

We now focus on two key statistical tests of knowledge restructuring; namely, changes in context sensitivity between Tests 1 and 2 (to test for initial knowledge restructuring), and context-sensitivity changes between Tests 3 and 4 (to test for strategy recovery). In the KP-first condition, there was a clear drop in context sensitivity between Tests 1 and 2 ($M = -.88$), reflecting knowledge restructuring from the KP to the CI strategy, $t(49) = -30.05, p < .001, r^2 = .95$. By contrast, in the CI-first condition, context sensitivity increased between Tests 1 and 2 ($M = .55$), reflecting restructuring from the CI to the KP strategy, $t(49) = 9.82$, $p < .001, r^2 = .66$. The patterns of restructuring between Tests 3 and 4 were consistent with recovery of people's original response strategy. In the KP-first condition, context sensitivity between Tests 3 and 4 increased ($M = .74$), as people restructured from the CI strategy in Test 3 to the KP strategy in Test 4, $t(49) = 12.80$, $p < .001, r^2 = .77$. In the CI-first condition, there was a reduction in context sensitivity ($M = -.62$), as people reverted back to the CI strategy, and away from knowledge partitioning, $t(49) = -10.57, p < .001, r^2 = .70$. To determine whether the knowledge restructuring between Tests 3 and 4 involved recovery of people's original response strategy, we correlated responses to the diagnostic stimuli (i.e., those that were sensitive to knowledge restructuring) from Tests 1 and 4 for the KP-first and CI-first conditions. In both cases, the correlations were very high, $r = .99$ (KP-first), $r = .98$ (CI-first), implying strategy recovery.

Structural Equation Modeling of Experiment 2

We investigated a number of structural equation models to explore the relationship between WMC and knowledge restructuring in the categorization task. Distributions of manifest variable data were either approximately Gaussian, or were log transformed to achieve normality, as discussed below.

Measurement Models. As in Experiment 1, we developed separate measurement models for the WMC battery and the categorization task, which we later combined into a structural model. The measurement model for WMC again included a single latent variable (WMC) and a freely-estimated pairwise correlation between the error terms associated with the OS and SS tasks. The model fit extremely well, $\chi^2(1) = 0.8$, $CFI = 1.0$, $RMSEA = 0.0$, 90% CI = [0.0, 0.255], $SRMR = .0132$. The loadings of the four manifest variables are shown in Table 7.

The measurement model for the categorization task involved 5 manifest variables. Three of the variables described training performance using the log transformed total number of errors: For Categorization Session 1, the two manifest variables described performance in blocks 1-3, and 4-6, respectively. For Categorization Session 2, a single variable described performance in blocks 7 and 8. The remaining two manifest variables corresponded to the absolute changes in the context-sensitivity measure within each testing session. Changes in context sensitivity reveal the extent of transition between the KP and CI strategies, and thus reflect the extent of knowledge restructuring within each session. Means and variances of the manifest variables are reported in Table 8.

We initially considered a two-factor measurement model, in which we associated one latent variable with the indices of learning, and the second latent variable with the representational attention measures (viz. the extent of knowledge restructuring in each session). On the basis of modification indices, a correlation between the error terms associated with training performance in the first and second halves of Categorization Session 1 was freely estimated. The two-factor model fit the data very well, $\chi^2(3) = 3.42$, $CFI = .997$, $RMSEA = .038$, 90% CI= [0.0, 0.178], $SRMR = .0275$, with all manifest variables loading in the expected way onto their latent variables. Constraining the correlation between the *Error* and *Knowledge Restructuring* latent variables to unity resulted in a significant decrement in fit, $\Delta\chi^2(1) = 4.08$, $p \approx .04$, justifying retention of a second factor in the model.

Structural Model. The measurement models for the WMC battery and the categorization task were combined to yield a three-factor structural model. The structural model fit the data well, $\chi^2(22) = 18.14$, $p = .70$, $CFI = 1.0$, $RMSEA = .0$, 90% CI = [0.0, 0.066], $SRMR = .0458$, and is presented in figure 9. The correlations among the three latent variables in Figure 9 are of the most interest. First, there is a strong negative correlation between the *Errors* latent variable and the *Knowledge Restructuring* latent variable, r =-.81, showing that more extensive restructuring was associated with more accurate learning. Second, *WMC* was negatively correlated with *Errors*, r =-.43, showing that WMC was related to fewer errors during learning. This relationship replicates the one we found in Experiment 1 between WMC and learning and it buttresses other work that has found a uniformly positive link between WMC and category learning (cf. Craig & Lewandowsky, in press; Lewandowsky, 2011; Lewandowsky et al., under review). Finally, there was a significant positive correlation between *WMC* and *Knowledge Restructuring*, $r = .36$, showing that higher WMC was related to a greater extent of restructuring. The latter result directly implicates an association between WMC and representational attention, showing that the two constructs share unique variance that is not accounted for by overall learning.

Individual variation in learning performance and the extent of knowledge restructuring were both shown to load onto latent variables associated with WMC. The pattern of results is consistent with the idea that the form of attention that is responsible for the coordination (and recoordination) of multiple categorization rules is related to the executive attention concept often invoked in working-memory research. By the same token, this form of attention differs from other manifestations of *learned* attention in category and association learning. We suggest that the knowledge restructuring in Experiment 2 was reliant on executive attention, whereas the highlighting effect from Experiment 1 was reliant on learned dimensional attention—only the former, but not the latter, shares unique variance with working memory capacity.

Implications

The key empirical contribution of Experiment 2 is the discovery that the extent of knowledge restructuring shares unique variance with WMC. The overall relationship is one of higher working memory capacity being associated with *greater* knowledge restructuring. That is, the extent to which people were able to successfully change categorization strategies was positively related to their working memory capacity. Because it is known that transitioning between the KP and CI strategies requires recoordination of partial category representations (Sewell & Lewandowsky, 2011), the correlation between the latent variables is plausibly attributable to an executive attention mechanism that enacts top-down selection of partial knowledge.

A clear implication for category learning is that the construct of "attention" must be nuanced to distinguish between feature-based or dimensional attention on the one hand and executive attention on the other. Although dimensional attention played a central role in performance in both experiments, only Experiment 2 required executive control over the coordination of multiple category representations.

General Discussion

We examined the relationship between WMC and two forms of attention in category/associative learning. Experiment 1 explored the relationship between WMC and *dimensional* attention using associative blocking and highlighting paradigms, which are widely thought to engage attentional factors (Kruschke et al., 2005). We found no evidence of any relationship between WMC and the magnitude of blocking and highlighting effects. Instead, we found that WMC related to overall learning performance (cf. Craig & Lewandowsky, in press; Lewandowsky, 2011; Lewandowsky et al., under review). Experiment 2 used a knowledge restructuring task that required *representational* attention to mediate changes in response strategies. In this case, we found that WMC related to both learning performance and the extent to which people could shift between response strategies. Taken together, the results imply that the relation between attention and WMC in categorization is determined by the need to *coordinate* multiple elements of partial knowledge (e.g., multiple categorization rules). In tasks that do not require coordination of multiple representations, WMC would only be expected to play a role in learning of stimulus-to-response associations (e.g., Lewandowsky, 2011). However, when the task requires coordination of multiple representations, people with higher WMC would be more effective at selectively accessing them, or alternatively, to be more effective at setting top-down attentional control parameters that determine how representational selection occurs on a trial-by-trial basis (e.g., as suggested by Erickson, 2008). Our results speak to a number of theoretical perspectives on attention, learning, and WMC. Before discussing these implications, we take up some potential limitations of the current studies.

Limitations

The principal limitation of the current experiments relates to the between-subjects nature of the comparisons between dimensional and representational attention. That is, different sets of people participated in the two experiments. This precludes comparison of the two modes of attention using an individual-differences approach. Pragmatic constraints prevented testing of the same individuals in all tasks.

A second potentially problematic aspect of the current results concerns the failure to observe a significant correlation between the magnitudes of the blocking and highlighting effects in Experiment 1. Our results stand in contrast to those of Kruschke et al. (2005) despite the fact that the tasks in our first study were largely identical to theirs.[8] Notwithstanding the necessary reluctance to accept a null result, our study appears unlikely to have been underpowered, given that our sample size was roughly three times that of Kruschke et al. (2005); namely 121 vs. 33. On the basis of modeling with EXIT (see Appendix A), we suggest that the absence of correlation likely reflects, at least in our study, differential involvement of associative and attentional mechanisms in blocking and highlighting. Theoretically, this result is in line with classical models that are able to explain the blocking effect without recourse to attention (Rescorla & Wagner, 1972). However, considered against the backdrop of more recent results that have shown attentionally mediated consequences of blocking (Beesley & Le Pelley, 2011; Kruschke, 2005b; Kruschke & Blair, 2000; Le Pelley et al., 2007), our results are somewhat puzzling. For now, we have good reason to accept that only highlighting involved attentional processes in our Experiment 1. The circumstances that determine whether a given example of blocking is attentional or associative in nature remain to be seen; the identification of such circumstances is a worthy target for future research.

Attention, Working Memory Capacity, and Category Learning

In category learning, attention has traditionally played a relatively circumscribed theoretical role, typically capturing the differential relevance of stimulus features (Kruschke, 2005a). For the most part, this is a parsimonious approach, as dimensional attention suffices to explain many category learning phenomena. More recently though, researchers investigating more complex category structures have shown that dimensional attention alone may not be enough to fully explain category learning (Aha & Goldstone, 1992; Denton, Kruschke, & Erickson, 2008; Erickson, 2008; Erickson & Kruschke, 1998, 2002a; Lewandowsky, Roberts, & Yang, 2006; Little & Lewandowsky, 2009; Sewell & Lewandowsky, 2011; Yang & Lewandowsky, 2003, 2004). In some cases, as in the ATRIUM model with its modular architecture, an additional "layer" of attention (i.e., representational attention) has been required to explain performance. The current study has further reinforced the need to distinguish between dimensional and representational (or executive) attention in category and associative learning. Whereas shifts in dimensional attention suffice to capture many aspects of learning that arise as a consequence of manipulating the dimensional relevance structure (e.g., Kruschke, 1996b), only shifts in representational attention are able to account for abrupt strategy shifts such as those related to knowledge restructuring (e.g., Sewell & Lewandowsky, 2011).

It is noteworthy that the representational attention shifts in Experiment 2 were controlled and volitional, in that participants had to deliberately change the way in which they coordinated rule use in order to comply with instructions. Although this is a distinctive feature of the task we used, and a driver of knowledge restructuring, there is nothing inherently volitional about shifting representational attention in ATRIUM, for example. We suggest that the deliberate control over the distribution of representational attention plays a particularly important role in relating knowledge restructuring and WMC. The need to selectively engage one modularized subset of knowledge at a time gels nicely with the fact that WMC relates to the ability to resist proactive interference (Engle, 2002; Engle et al., 1999; Kane et al., 2001; Oberauer & Kliegl, 2001). That people in our experiment were able to successfully reinstate their original response strategy in Test 4 of the category learning task is further suggestive of this protective function of executive attention. The alignment of executive attention with WMC has clear implications for perspectives on category learning that foreground the role of working memory.

Recently, several category learning theorists have hypothesized that working memory might be differentially involved in different types of categorization tasks (Ashby & Maddox, 2005). These authors have suggested that rule-based tasks—such as those we used here—tax working memory, whereas other so-called information-integration tasks do not (Ashby & O'Brien, 2005; see Newell, Dunn, & Kalish, 2011, for a contrasting perspective). Information-integration tasks are not solvable by a verbalizable rule and require the integration of two or more aspects of the stimulus at a pre-decisional stage. In contrast to that expectation, our relevant work to date has produced quite stable and reproducible results: Across six experiments involving over 800 participants in total, WMC has been found to be associated with overall levels of learning, regardless of whether tasks place differential demands on dimensional attention (Craig & Lewandowsky, in press; Lewandowsky, 2011), or are rule-based or require information-integration (Lewandowsky et al., under review). We attribute the stability of our findings to our latent variable approach to measuring WMC, which permits a more robust assessment of WMC that is free of measurement error and avoids problems arising from the substantial task-specific variance associated with individual working memory tasks (Lewandowsky et al., 2010).

Top-down vs. Bottom-up Processing

Our provisional alignment of executive and representational attention emphasizes the contribution that controlled, top-down, processes make to category learning. The emphasis on volitional processing is underscored by the fact that knowledge restructuring was elicited by a deliberate shift of representational attention. In complex categorization tasks such as those involving task partitioning or knowledge restructuring, the controlled coordination of different components of partial knowledge is critical. For example, the knowledge restructuring observed in Experiment 2 required systematic changes in the way rule representations were selectively accessed over the course of the experiment. The final session of Experiment 2 in particular required actively applying one response strategy whilst passively maintaining the alternative strategy. The simultaneous selection and suppression of subsets of strategic information bears close resemblance to notions of executive attention (e.g., Engle & Kane, 2004; Kane et al., 2007). The fact that WMC was strongly associated with the extent of knowledge restructuring is quite consistent with the idea of executive attention playing a key role in recoordination in particular, and knowledge restructuring more generally.

The link between representational attention and WMC is readily contrasted with the absence of any relationship between WMC and dimensional attention (e.g., Experiment 1, and Lewandowsky, 2011). The latter is suggestive of a contribution of stimulus-driven, bottom-up, processes in categorization tasks that do not require coordination of knowledge or executive control. The idea that highlighting and more elaborate blocking designs may involve bottom-up factors is supported by the finding that manipulating the salience of the blocked cue can greatly attenuate the blocking effect (Denton & Kruschke, 2006). Similarly, Lamberts and Kent (2007) argued that highlighting is unlikely to be mediated by strategic, top-down factors, such as explicit hypothesis testing because the effect persists even under severe time pressure to respond (e.g., within 300 and 500 ms). By a similar token, there is much evidence to suggest that bottom-up attentional factors are unrelated to WMC. Kane et al. (2001) examined the relationship between working memory span and performance in prosaccade and antisaccade tasks where people, respectively, had to orient attention toward or away from a highly salient peripheral cue (e.g., Posner, 1980). Whereas the prosaccade task involved exogenous (bottom-up) reflexive orienting, the antisaccade task involved endogenous (top-down) controlled orienting. Only performance on the *antisaccade* task was related to working memory span.

Alternative views of working memory

Although we have interpreted our results within the framework of an executive-attention view of working memory, and although our results mesh well with that notion, we do not selectively endorse this view over other theoretical approaches to working memory. For example, Oberauer and colleagues (e.g., Oberauer, Süß, Wilhelm, & Sander, 2007) have developed a tripartite approach to working memory that 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 (category) learning is thought to involve transfer of information from the direct-access region to long-term memory. 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 theoretical development of a process model relating WMC to long-term category learning. As a step in that direction, Lewandowsky

(2011) showed that WMC can map into additional "rehearsals" within ALCOVE, suggesting that people with higher WMC are better able to retain bound representations of stimuli in working memory for successive strengthening of their long-term connections with the response options. It remains for future research to explore how the binding notion favored by Oberauer and colleagues maps into the executive-attention notion required to explain the knowledge restructuring results of Experiment 2.

Future Directions

Although this paper has addressed a number of issues on the emerging relationship between working memory and categorization, our work has also opened up a number of theoretical questions that remain unanswered. We thus briefly offer suggestions for future research. With regards to theories of learning and attention, our failure to find an attentional locus of the blocking effect (see Appendix A for full details) can be contrasted with a variety of findings that have found attentional consequences of blocking (e.g., Beesley & Le Pelley, 2011; Kruschke, 2005b; Kruschke & Blair, 2000). A key factor seems to be the overall complexity of the blocking design. Studies that have found evidence for attentional effects of blocking have usually used more complicated designs that involve multiple testing phases, and larger sets of stimulus cues. Clearly, a more comprehensive understanding of the factors that determine whether a blocking effect is attentionally mediated is needed.

Turning to the relationship between WMC and categorization, our study found evidence that WMC was related to category learning in both experiments. In addition, WMC predicted knowledge restructuring. The abrupt changes in performance in our knowledge restructuring task relate closely to other tasks that require set-shifting of various kinds. To date, the evidence that set-shifting relates to WMC has been surprisingly weak. Oberauer et al. (2003) found only weak relationships between set-shifting ability and performance on a number of standard working memory tasks (e.g., reading span). Similarly, Miyake, Friedman, Emerson, Witzki, and Howerter (2000) found only weak relationships between set-shifting and operation span performance. In reconciling our results with those of Miyake et al. (2000) and Oberauer et al. (2003), it is useful to refer to the distinction made by Cools, Ivry, and D'Esposito (2006) between shifts that involve concrete aspects of a task (e.g., responding on the basis of different stimulus features on different trials) vs. those that involve abstract aspects of a task (e.g., responding on the basis of a different rule from a previous trial). The former type of shift involves a change in the stimulus to be responded to. That is, a different stimulus will drive responding on different trials, but the response rule might stay the same. The latter type of shift involves applying different response rules to the same stimulus at different times. The tasks used by Miyake et al. (2000) and Oberauer et al. (2003) involved one or the other type of shift. By contrast, our Experiment 2 required shifting between qualitatively different ways of coordinating common categorization rules: Participants had to shift between rule sets on a trial-by-trial basis (i.e., shifting between bar position and rectangle height rules), but also had to shift between response strategies that involved different stimulus features (i.e., using context or bar position to select a rule). Resolution of the sorts of set-shifting that do and do not relate to WMC awaits further research.

Another aspect of performance in our knowledge restructuring task was the "gating" function played by different stimulus dimensions under different categorization strategies. For example, the KP strategy required people to categorize stimuli on the basis of *either* the rectangle height or bar position dimensions depending on the context dimension. Thus, certain stimulus dimensions were not always directly relevant for categorization, but were nevertheless useful in that they indicated *other* dimensions that were directly relevant for categorization (cf. Blair, Watson, Walshe, & Maj, 2009). Given that WMC was found to correlate with people's ability to change the dimension they used to gate rule use, it is possible that WMC might only underpin categorization strategies that involve such dimensional gating. However, Craig and Lewandowsky (in press) have explored strategy use in a number of categorization tasks and found no relationship between WMC and the categorization strategy people ultimately used (though categorization accuracy was related to WMC, much as we report here, regardless of which strategy people ultimately chose). Clarification of how categorization strategy, coordination of representations, and WMC interrelate is, in our view, a very worthy target for future research.

Conclusions

Attention is a central but heterogeneous theoretical construct in the areas of category learning and working memory. Despite this apparent overlap, it has remained unclear whether constructs included under the rubric of attention generalize across research domains. Recent studies investigating the relationship between WMC and dimensional attention in categorization have been unable to find a reliable link (e.g., Craig & Lewandowsky, in press; Lewandowsky, 2011; Lewandowsky et al., under review), suggesting differences beteween attentional constructs in category learning and working memory theorizing. We have shown that WMC is related to a different form of attention in categorization that controls the coordination of multiple category representations. The latter property is only likely to manifest in more complex categorization tasks that involve top-down selection of partial knowledge.

References

Aha, D. W., & Goldstone, R. L. (1992). Concept learning and flexible weighting. In J. K. Kruschke (Ed.), *Proceedings of the 14th annual conference of the cognitive science society* (p. 534-539). Hillsdale, NJ: Erlbaum.

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.

Beesley, T., & Le Pelley, M. E. (2011). The influence of blocking on overt attention and associability in human learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 114-120.

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *35*, 1196-1206.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.

Charles, E. P. (2005). The correction for attenuation due to measurement error: Clarifying concepts and creating confidence sets. *Psychological Methods*, *10*, 206-226.

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.

Cools, R., Ivry, R. B., & D'Esposito, M. (2006). The human striatum is necessary for responding to changes in stimulus relevance. *Journal of Cognitive Neuroscience*, *18*, 1973-1983.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, *24*, 87-185.

Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.

Craig, S., & Lewandowsky, S. (in press). Whichever way you choose to categorize, working memory helps you learn. *Quarterly Journal of Experimental Psychology*.

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

Denton, S. E., & Kruschke, J. K. (2006). Attention and salience in associative blocking. *Learning & Behavior*, *34*, 285-304.

Denton, S. E., Kruschke, J. K., & Erickson, M. A. (2008). Rule-based extrapolation: A continuing challenge for exemplar models. *Psychonomic Bulletin & Review*, *15*, 780-786.

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, & Cognition*, *36*, 170-189.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19-23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, p. 145-199). Elsevier.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309-331.

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. (2002a). Multiple representations in inductive category learning: Evidence of stimulus- and task-dependent representation. (Unpublished manuscript.)

Erickson, M. A., & Kruschke, J. K. (2002b). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, *9*, 160-168.

Fan, X. (2003). Two approaches for correcting correlation attenuation cause by measurement error: Implications for research practice. *Educational and Psychological Measurement*, *63*, 915-930.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227-247.

Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General*, *139*, 319-340.

Johansen, M. K., Fouquet, N., & Shanks, D. R. (2010). Featural selective attention, exemplar representation, and the inverse base-rate effect. *Psychonomic Bulletin & Review*, *17*, 637-643.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Kalish, M. L. (2001). An inverse base rate effect with continuously valued stimuli. *Memory & Cognition*, *29*, 587-597.

Kalish, M. L., & Kruschke, J. K. (2000). The role of attention shifts in the categorization of continuous dimensioned stimuli. *Psychological Research*, *64*, 105-116.

Kalish, M. L., Lewandowsky, S., & Davies, M. (2005). Error-driven knowledge restructuring in categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 846-861.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation* (p. 9-33). Coral Gables, FL: University of Miami Press.

Kamin, L. J. (1969). Predictability, surprise, attention, and condtioning. In R. M. Church & B. A. Campbell (Eds.), *Punishment and aversive behavior* (p. 279-296). New York: Appleton-Century-Crofts.

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, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (p. 21-48). Oxford: Oxford University Press.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70.

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

Kruschke, J. K. (1996a). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *22*, 3-26.

Kruschke, J. K. (1996b). Dimensional relevance shifts in category learning. *Connection Science*, *8*, 225-247.

Kruschke, J. K. (2001a). The inverse base-rate effect is not explained by eliminative inference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *27*, 1385-1400.

Kruschke, J. K. (2001b). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812-863.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*, 171-175.

Kruschke, J. K. (2005a). Category learning. In K. Lamberts & R. L. Goldstone (Eds.), *The handbook of cognition* (p. 183-201). London: Sage.

Kruschke, J. K. (2005b). Learning involves attention. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (p. 113-140). Hove, East Sussex, UK: Psychology Press.

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (p. 267-301). Cambridge: Cambridge University Press.

Kruschke, J. K. (2009). Highlighting: A canonical experiment. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, p. 153-185).

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*, 636-645.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, & 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 & Cognition*, *31*, 830-845.

Lamberts, K., & Kent, C. (2007). No evidence for rule-based processing in the inverse base-rate effect. *Memory & Cognition*, *35*, 2097-2105.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental*

Psychology, *57B*, 193-243.

Le Pelley, M. E., Beesley, T., & Suret, M. B. (2007). Blocking of human causal learning involves learned changes in stimulus processing. *Quarterly Journal of Experimental Psychology*, *60*, 1468-1476.

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *37*, 720-738.

Lewandowsky, S., Kalish, M., & Griffiths, T. L. (2000). Competing strategies in categorization: Expediency and resistance to knowledge restructuring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 1666-1684.

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.

Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (under review). Working memory does not dissociate between different perceptual categorization tasks.

Little, D. R., & Lewandowsky, S. (2009). Beyond nonutilization: Irrelevant cues can gate learning in probabilistic categorization. *Journal of Experimental Psychology: Human Perception & Performance*, *35*, 530-550.

Little, D. R., Lewandowsky, S., & Heit, E. (2006). Ad hoc category restructuring. *Memory & Cognition*, *34*, 1398-1413.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376-400.

Logan, G. D. (2004). Cumulative progress in formal theories of attention. *Annual Review of Psychology*, *55*, 207-234.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309-332.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276-298.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General*, *117*, 68-85.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: fact or fantasy? In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, p. 167-215).

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *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: General*, *115*, 39-57.

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

Oberauer, K. (1993). Die koordination kognitiver operationen: Eine studie über die beziehung zwischen intelligenz und "working memory" (the coordination of cognitive operations: A study on the relation of intelligence and "working memory"). *Zeitschrift für psychologie*, *201*, 57-84.

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, p. 45-100).

Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology*, *13*, 187-215.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittman, W. W. (2000). Working memory capacity — facets of a cognitive ability construct. *Personality & 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* (p. 49-75). Oxford: Oxford University Press.

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

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

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3-25.

Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1-41.

Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 811-829.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black &

W. F. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (p. 64-99). New York: Appleton-Century-Crofts.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology & Aging*, *6*, 118-127.

Sewell, D. K., & Lewandowsky, S. (2011). Restructuring partitioned knowledge: The role of recoordination in category learning. *Cognitive Psychology*, *62*, 81-122.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology*, *37*, 1-21.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General & Applied*, *75*, 1-42.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 3-27.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72-101.

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.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*, 282-299.

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

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

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

Appendix A

Fit of EXIT to Experiment 1 Data

We investigated the lack of correlation between blocking and highlighting effects via computational modeling with EXIT (Kruschke, 2001b). EXIT comprises an associative network that connects cues to outcomes. The model also incorporates a complex attentional mechanism that combines rapid attention shifts upon presentation of a stimulus with long-term attentional learning across trials. Our guiding question was why the blocking-highlighting correlation was absent from our data. One possibility is that our participants differentially relied on attention across the two tasks. Kruschke et al. (2005) showed via simulations with EXIT that the model could generate a positive correlation between blocking and highlighting effects if its attentional mechanisms were engaged in both tasks. However, their Figure 2 reveals that within the model, dimensional attention was only actually required to generate a highlighting effect. By contrast, blocking effects, although increased by attention, arise as a more fundamental consequence of the model's error-driven learning mechanism. This is theoretically unsurprising, as EXIT contains a form of the Rescorla and Wagner (1972) model as a special case (Kruschke, 2001b), and that model suffices to produce blocking effects without dimensional attention, through error-driven learning only. It follows that if our participants performed the blocking task by exclusively relying on an associative learning mechanism, but recruited both attentional and associative mechanisms to perform the highlighting task, no correlation between blocking and highlighting effects would be expected. We next provide a brief overview of EXIT (Kruschke, 2001b, 2001a, provides complete details).

EXIT combines error-driven association learning with rapid shifts of attention in a connectionist framework (cf. Kruschke & Johansen, 1999). As discussed by Kruschke (2001b), when EXIT's attentional system is active, the model essentially includes

Mackintosh's (1975) as a special case; similarly, when the attentional system is inactive, EXIT reduces to a form of the Rescorla and Wagner (1972) model. On a given trial, the model responds on the basis of learned associations between cues and response outcomes. When corrective feedback is encountered, attention is rapidly shifted away from cues that generated prediction error and onto cues that lead to more accurate performance. The extent of attention shifting is governed by a shift parameter, λ_g. After an attention shift has occurred, association weights are updated such that learning of cue-outcome associations is focused on cues that will maximize performance. The extent of association learning is governed by a learning rate parameter, λ_w. In addition to its rapid attention shift mechanism, EXIT also incorporates an attention *learning* mechanism that is tied to exemplar memory (cf. Kruschke, 1992). The rationale is that different attentional distributions may be suitable for different stimulus configurations (e.g., Aha & Goldstone, 1992). Thus, EXIT attempts to learn the post-shift distribution of dimensional attention so it can be applied when the stimulus is next encountered. The extent to which the model learns the shifted distribution of dimensional attention is governed by an attention learning parameter, λ_x, and the degree to which the shifted attention distribution is tied exclusively to a given stimulus is determined by an exemplar specificity parameter, c.

To summarize, EXIT has three parameters that relate to dimensional attention: The attention shift parameter, λ_g, the attention learning parameter, λ_x, and the exemplar specificity parameter, c. Cue-association learning is controlled by the λ_w parameter. In addition to these parameters, EXIT also includes a normalizing constant used when computing attention weights, P, and a decision parameter, ϕ, which converts response node activations to choice probabilities.

To capture our assumptions that the role of attentional and associative mechanisms may have differed between the blocking and highlighting tasks, we allowed the three attention parameters (λ_g, λ_x, and c) and the association learning parameter (λ_w) to assume task-specific values. We fit EXIT directly to the choice probability data for each unique combination of cues presented at test. Parameters were estimated by minimizing the (negative) multinomial log likelihood statistic, -ln L =-$\sum_{i=1}^{20}\sum_{j=1}^{4} d_{ij}\ln(p_{ij})$. The outer summation over i indexes the different cue combinations presented at test. The inner summation over j indexes the 4 response outcomes, and p_{ij} and d_{ij} correspond, respectively, to the predicted proportion and observed frequency of outcome j responses for cue combination i. Best fitting parameters are presented in Table A1.

EXIT provided an excellent fit to the combined blocking and highlighting data, $RMSD = .0263$. The pattern of parameter values are readily interpreted: Attention was clearly not required to fit the blocking data, whereas association learning was; best fitting values for $\lambda_{g,B}$, $\lambda_{x,B}$, and c_B were all approximately 0, whereas $\lambda_{w,B}$ was 1.98. It is noteworthy that under this set of parameter values, when EXIT's attentional mechanisms are inactive, the model closely approximates the Rescorla and Wagner (1972) model (see Kruschke, 2001b, for discussion). To fit the highlighting data, by contrast, EXIT required both attentional and associative mechanisms to be active; $\lambda_{g,H}$, $\lambda_{x,H}$, c_H, and λ_{w_H} were all greater than 0. To confirm that this level of theoretical flexibility was necessary to explain the data, we also fit a restricted version of EXIT that did not allow attention and association parameters to vary across tasks. Although this model was able to reproduce all qualitative patterns in the data, quantitatively the restricted model fit significantly worse than the more general model discussed above, -ln $L = 4930.69$, $RMSD = .0878$, $\Delta\chi^2(4) \approx$ -2 $\ln L_{\text{restricted}} + 2\ln L_{\text{general}} = 601.06, p < .05$. Best fitting parameters for the restricted model were: $c = 1.51$, $\phi = 4.84$, $P = 1$, $\lambda_g = .11$, $\lambda_x = .0002$, $\lambda_w = .64$. Taken together, the modeling shows that, at least in our experiment, the blocking and highlighting effects arose from different mechanisms; the former were associative, the latter, attentional.

Appendix B

Extended Analysis of Experiment 1

Lewandowsky (2011) recently reported a robust correlation between WMC on the one hand and associative learning rate on the other. In light of Lewandowsky's result, and given that our blocking effect was driven by associative learning factors, the failure of blocking performance in Experiment 1 to load onto the *Errors* latent variable is striking. We suggest that the reason blocking failed to load onto the latent variable is because the conventional measure of blocking used by both Kruschke et al. (2005) and us, is insensitive to important aspects of the associative history of the stimulus. We show in this extended analysis that if the blocking measure is modified to account for associative history, performance loads onto the latent variable in the expected manner.

Recall that in Experiment 1, a pair of cues was presented on each test trial in the blocking task. Each cue had an associative history with a distinct outcome. For example, the cues comprising the test compound B1.D1, were trained with outcomes X1 and Y1 respectively. People had to respond to those items by choosing among 4 possible outcomes: X1, X2, Y1, and Y2. Thus, there were always 2 outcomes that were *relevant* to the cues (i.e., outcomes trained with the cues) and 2 outcomes that were *irrelevant* to the cues (i.e., outcomes never trained with the cues). Blocking was measured by examining the relative rates of generating *relevant* response outcomes. For example, for the test item B1.D1, only outcomes X1 and Y1 were considered, whereas X2 and Y2 responses were effectively omitted from the calculation. If excluding *irrelevant* responses obscures the associative nature of the effect, a modified performance index that incorporates this information would be expected to relate to WMC, and thus load on the *Error* latent variable.

We constructed alternative performance indices that were similar to the *AmCc* and *DmBc* measures reported in the main text. However, instead of taking the difference between the number of relevant X and Y responses, we summed the number of *relevant* responses before dividing by the number of test items, thus yielding an index of the number of *relevant* responses made to test stimuli. Because the modified measures effectively assess the combined associative strength of all stimulus cues in the display with their trained outcomes relative to untrained outcomes, they do not assess attentional effects like the standard measures do. We summed the modified component measures to get an overall relevance index for blocking and highlighting, respectively, which we used as manifest variables to replace Blocking and Highlighting (cf. Figure 3). This revised SEM model, with the correlation between error terms of the two relevance measures freely estimated, fit well, $\chi^2(32) = 44.85$; $CFI = .953$; $RMSEA = 0.058$, 90% CI = [0.0, 0.095], $SRMR = .0598$. The loadings of the blocking ($p < .0001$) and highlighting ($p = .061$) relevance measures on the *Error* latent variables imply that the failure of the conventional blocking measure to load onto the *Error* latent variable arose because the measure ignored an important associative aspect of performance. When the measure is augmented to take this associative aspect in account, it loads onto the latent variable as expected. Thus, the apparent discrepancy between our Experiment 1 results and those of Lewandowsky (2011) can be attributed to a limitation in the way blocking is conventionally measured.

Appendix C

Correlation Matrices for all Manifest Variables Used in the Structural Equation Models in both Experiments

[Tables B1 and B2 to go here]

Author Note

This research was supported by a grant from the Australian Research Council to the second author and Gilles Gignac and by an Australian Professorial Fellowship to the second author. We gratefully acknowledge the support of Philip Smith during completion of this project. We thank Charles Hanich for assistance with data collection, and four anonymous reviewers for their comments. Address correspondence to the first author at The University of Melbourne, Melbourne, VIC 3010, Australia. Electronic mail may be sent to the first author at dsewell@unimelb.edu.au, or the second author at stephan.lewandowsky@uwa.edu.au. Web pages:
http://www.psych.unimelb.edu.au/people/staff/SewellD.html (DKS);
http://www.cogsciwa.com. (SL)

Notes

[1] For convenience, we discuss selection of a category representation as if it were an all-or-none process, but note that more flexible weighting schemes are used in practice. We further elaborate on the distinction between dimensional and representational attention when we introduce Experiment 2.

[2] Although "attention" in these kinds of tasks can be decomposed into dimensional attention on the one hand and cue- or feature-based attention on the other (e.g., Kalish, 2001; Johansen et al., 2010), we couch our discussion of blocking and highlighting effects in terms of dimensional attention. It is noteworthy though that when stimuli are comprised of cues that can be present or absent (as in Experiment 1, along with most investigations of blocking and highlighting), stimulus dimensions become indistinguishable from stimulus features.

[3] The highlighting effect was first reported by Medin and Edelson (1988) as an *inverse base rate effect*, so-called because their study involved only a single training phase with the base rates of stimuli analogous to I.pE $\rightarrow$ E and I.pL $\rightarrow$ L differing according to a 3:1 ratio. It is now known that unequal base rates are not required to produce the effect (Kruschke, 2009), and hence the more appropriate term "highlighting" is used to refer to the phenomenon.

[4] For the model to be estimable, the residual variance for MU had to be fixed at 0. This proved unnecessary for the structural model, and we are therefore not overly concerned about this constraint.

[5] We also investigated models that included the "pure" measures of blocking and highlighting as manifest variables (i.e., $DmBc$ and $pLmpEc$). These component measures, like the other measures we examined in the main text, failed to load onto the latent variable.

[6] Technically, there are multiple CI strategies. Rule selection could be determined by position along the y axis, or through some combination of x and y positions. We describe the strategy in relation to the x axis because we experimentally controlled strategy use via direct instructions.

[7] The category learning data from the first 48 participants were analyzed and modeled in the aggregate and reported elsewhere (Sewell & Lewandowsky, 2011). That initial report did not include any WMC results or individual-differences analysis and also did not include the data of the remaining 58 participants.

[8] One procedural factor that may have contributed to the discrepancy was that we trained people for a fixed number of trials in Experiment 1. By contrast, Kruschke et al. (2005) trained people to an accuracy criterion. Although use of accuracy criteria to assess task mastery can be fraught with problems (e.g., Tharp & Pickering, 2009), the fact that we were able to reproduce blocking and highlighting effects at asymptotic levels of performance allays these concerns.

Table 1

Typical blocking and highlighting designs. Typical responses to test items are shown in parentheses.

Phase	Blocking		Highlighting	
Early	A→X	F→Y	I.pE→E	
Late	A.B→X	C.D→Y	I.pE→E	I.pL→L
Test	B.D→? (Y)	B.C→? (Y)	pE.pL→? (L)	
	A.C→? (X)	A.D→? (X)	I→? (E)	

Table 2

Details of the blocking design used in Experiment 1. The ordering of cues in the table reflects their left-right positioning on the screen in the experiment. The "_" symbol denotes a cue position that was unoccupied. Stimuli unique to the test phase are divided into two classes: B.(C/D)→? and A.(C/D)→?; see text for details.

Phase	Trial Items				Duration
Early	A1._→X1	_.A1→X1	F1._→Y1	_.F1→Y1	10 × 8 trial blocks
	A2._→X2	_.A2→X2	F2._→Y2	_.F2→Y2	
Late	A1.B1→X1	B1.A1→X1	C1.D1→Y1	D1.C1→Y1	10 × 8 trial blocks
	A2.B2→X2	B2.A2→X2	C2.D2→Y2	D2.C2→Y2	
Test	Training Stimuli Shown at Test (presented twice each)				
	A1.B1→X1	B1.A1→X1	C1.D1→Y1	D1.C1→Y1	
	A2.B2→X2	B2.A2→X2	C2.D2→Y2	D2.C2→Y2	
	Test Stimuli: Class B.(C/D)→?				
	D1.B1→?	B1.D1→?	D2.B2→?	B2.D2→?	
	D2.B1→?	B1.D2→?	D1.B2→?	B2.D1→?	48 trials
	C1.B1→?	B1.C1→?	C2.B2→?	B2.C2→?	
	C2.B1→?	B1.C2→?	C1.B2→?	B2.C1→?	
	Test Stimuli: Class A.(C/D)→?				
	A1.C1→?	C1.A1→?	A2.C2→?	C2.A2→?	
	A1.C2→?	C2.A1→?	A2.C1→?	C1.A2→?	
	A1.D1→?	D1.A1→?	A2.D2→?	D2.A2→?	
	A1.D2→?	D2.A1→?	A2.D1→?	D1.A2→?	

Table 3

Details of the highlighting design used in Experiment 1. The ordering of cues in the table reflects their left-right positioning on the screen in the experiment. The "_" symbol denotes a cue position that was unoccupied. Stimuli unique to the test phase are divided into two classes: I._→? and pE.pL→?; see text for details.

Phase	Trial Items				Duration
Early	I1.pE1→E1	pE1.I1→E1			3 × 8 trial blocks
	I2.pE2→E2	pE2.I2→E2			
Late(1)	I.pE shown *3 times each*:		I.pL shown *1 time each*:		3 × 16 trial blocks
	I1.pE1→E1	pE1.I1→E1	I1.pL1→L1	pL1.I1→L1	
	I2.pE2→E2	pE2.I2→E2	I2.pL2→L2	pL2.I2→L2	
Late(2)	I.pE shown *1 time each*:		I.pL shown *3 times each*:		6 × 16 trial blocks
	I1.pE1→E1	pE1.I1→E1	I1.pL1→L1	pL1.I1→L1	
	I2.pE2→E2	pE2.I2→E2	I2.pL2→L2	pL2.I2→L2	
Test	Training Stimuli Shown at Test (shown *1 time each*)				24 trials
	I1.pE1→E1	pE1.I1→E1	I1.pL1→L1	pL1.I1→L1	
	I2.pE2→E2	pE2.I2→E2	I2.pL2→L2	pL2.I2→L2	
	Test Stimuli shown *2 times each*				
	Class I._→?		Class pE.pL→?		
	I1._→?	_.I1→?	pE1.pL1→?	pL1.pE1→?	
	I2._→?	_.I2→?	pE2.pL2→?	pL2.pE2→?	

Table 4

Performance on the Working Memory Tasks in Experiment 1

Measure	MU	OS	OS_{pt}	SS	SS_{pt}	SSTM
Mean	0.59	0.71	0.91	0.68	0.91	0.86
SD	0.19	0.14	0.08	0.16	0.06	0.06
Maximum	1.00	1.00	1.00	1.00	1.00	1.00
Minimum	0.16	0.16	0.16	0.16	0.16	0.16
Kurtosis	2.26	3.14	28.87	3.78	4.76	3.38
Skewness	-0.07	-0.60	-3.97	-0.82	-1.18	-0.48
SEM weights	1.00	0.56		0.55		0.44

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

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

Table 5

Cronbach's αs for overall indices of blocking and highlighting, and stimulus subsets comprising their component measures.

	Blocking	B.(C/D)	A.(C/D)	Highlighting	pE.pL	I._
Cronbach's α	.76	.64	.64	.57	.52	.53

Table 6

Means, standard deviations, skewness, and kurtosis for manifest variables from the blocking and highlighting tasks. All error measures were log transformed.

	Mean	SD	Skewness	Kurtosis
Blocking Early Errors	1.26	0.71	0.32	0.63
Blocking Late Errors	0.79	0.69	.51	-0.37
Highlighting Early Errors	0.56	0.50	0.57	0.40
Highlighting Middle & Late Errors	1.29	0.82	0.32	-0.001
Blocking Effect	0.62	0.63	0.34	-0.79
Highlighting Effect	0.67	0.68	0.21	-0.48

Table 7

Performance on the Working Memory Tasks in Experiment 2

Measure	MU	OS	OS_{pt}	SS	SS_{pt}	SSTM
Mean	0.62	0.74	0.93	0.71	0.92	0.87
SD	0.16	0.11	0.04	0.13	0.05	0.04
Maximum	0.99	0.99	0.99	0.99	0.99	0.99
Minimum	0.19	0.19	0.19	0.19	0.19	0.19
Kurtosis	2.38	2.81	6.48	4.83	4.64	2.50
Skewness	-0.31	-0.30	-1.31	-0.88	-1.25	-0.46
SEM weights	0.71	0.54		0.56		0.56

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

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

Table 8

Means, standard deviations, skewness, and kurtosis for manifest variables from the knowledge restructuring task. All error measures were log transformed.

	Mean	SD	Skewness	Kurtosis
Session 1 Early Errors	2.38	0.88	-0.03	-0.77
Session 1 Late Errors	1.03	1.12	.92	-0.21
Session 2 Errors	1.36	0.80	0.29	0.04
Session 1 Extent of Restructuring	0.74	0.29	-1.23	0.71
Session 2 Extent of Restructuring	0.73	0.32	-1.25	0.45

Table A1

Summary of fit statistics and best fitting parameters of EXIT fit to data from Experiment 1. Parameter subscripts identify best fitting values for the blocking (B) and highlighting (H) tasks. Parameters correspond to exemplar specificity (c), response decisiveness (ϕ), attention capacity (P), attention shift rate (λ_g), attention learning rate (λ_x), and association learning rate (λ_w). See Kruschke (2001) for details.

-ln L	RMSD	c_B	c_H	ϕ	P	$\lambda_{g,B}$	$\lambda_{g,H}$	$\lambda_{x,B}$	$\lambda_{x,H}$	$\lambda_{w,B}$	$\lambda_{w,H}$
4630.16	.0263	0	.74	6.78	1.88	0	.08	0	9.85	1.98	.82

Table C1

Correlations between all WMC tasks and learning and attentional manifest variables in Experiment 1

	WMC tasks				Learning				Attention	
	MU	OS	SS	SSTM	B(early)	B(late)	H(early)	H(midlate)	B	H
MU	—									
OS	0.56	—								
SS	0.55	0.74	—							
SSTM	0.44	0.24	0.23	—						
B(early)	-0.33	-0.21	-0.27	-0.22	—					
B(late)	-0.18	-0.12	-0.19	-0.15	0.42	—				
H(early)	0.00	0.00	-0.14	-0.00	0.27	0.26	—			
H(midlate)	-0.31	-0.05	-0.07	-0.19	0.39	0.33	0.05	—		
B	0.05	-0.01	-0.02	0.05	-0.09	0.01	0.00	-0.13	—	
H	-0.06	-0.05	-0.09	0.05	-0.09	-0.05	0.07	-0.07	0.04	—

Legend. WMC tasks are as in Table 4. B=blocking; H=highlighting.

Table C2

Correlations between all WMC tasks and category learning and knowledge restructuring manifest variables in Experiment 2

	WMC tasks				Category Learning			Knowledge Restructuring	
	MU	OS	SS	SSTM	S1(early)	S1(late)	S2(all)	KR1	KR2
MU	—								
OS	0.37	—							
SS	0.41	0.66	—						
SSTM	0.40	0.33	0.29	—					
S1(early)	-0.28	-0.23	-0.14	-0.25	—				
S1(late)	-0.20	-0.14	-0.06	-0.11	0.67	—			
S2(all)	-0.22	-0.26	-0.15	-0.15	0.44	0.52	—		
KR1	0.26	0.18	0.18	0.18	-0.33	-0.36	-0.38	—	
KR2	0.12	0.21	0.13	0.06	-0.35	-0.32	-0.53	0.48	—

*Legend.*WMC tasks are as in Table 7. S1=Session 1 categorization performance; S2=Session 2 categorization performance; KR1=extent of knowledge restructuring at end of Session 1; KR2=extent of knowledge restructuring at end of Session 2.

Figure Captions

Figure 1. Schematic overview of stages of processing involved in categorization. Points at which attentional mechanisms might operate are drawn as dashed boxes; see text for details. Perceptual processes provide raw inputs along a number of stimulus dimensions. Dimensional attention then weights the input dimensions. When multiple category representations are available, representational attention is used to select a suitable category representation. The selected representation is then activated by the weighted dimensional input, usually on the basis of similarity. A category decision is then made on the basis of the activation pattern, resulting in a response.

Figure 2. Schematic of an example blocking trial. Stimulus words are presented toward the top of a computer screen. The four response options are presented toward the bottom of the screen. Participants indicated their response via mouse click.

Figure 3. Structural model for Experiment 1. Standardized estimates and all statistically significant paths are presented in bold.

Figure 4. Illustration of the distinction between dimensional and representational attention. Attentional effects are shown as boldfaced associative connections. The top panel shows a case involving only dimensional attention. Connections from input dimensions D1 and D2 to the exemplar nodes (triangles) are selectively enhanced by being weighted more heavily than D3. The bottom panel shows a case involving both dimensional and representational attention. The pattern of connections from inputs to exemplars from the top panel are reproduced in the left-hand side of the bottom panel. The effects of representational attention are shown via the enhanced associations from individual exemplars to the various representational mappings. Exemplars 1 and 2 gate access to representation R1, whereas Exemplars 3 and 4 gate access to representation R2.

Each representation summarizes a stimulus-response mapping (e.g., an associative matrix, such as the one shown in the shaded rectangle in the top panel).

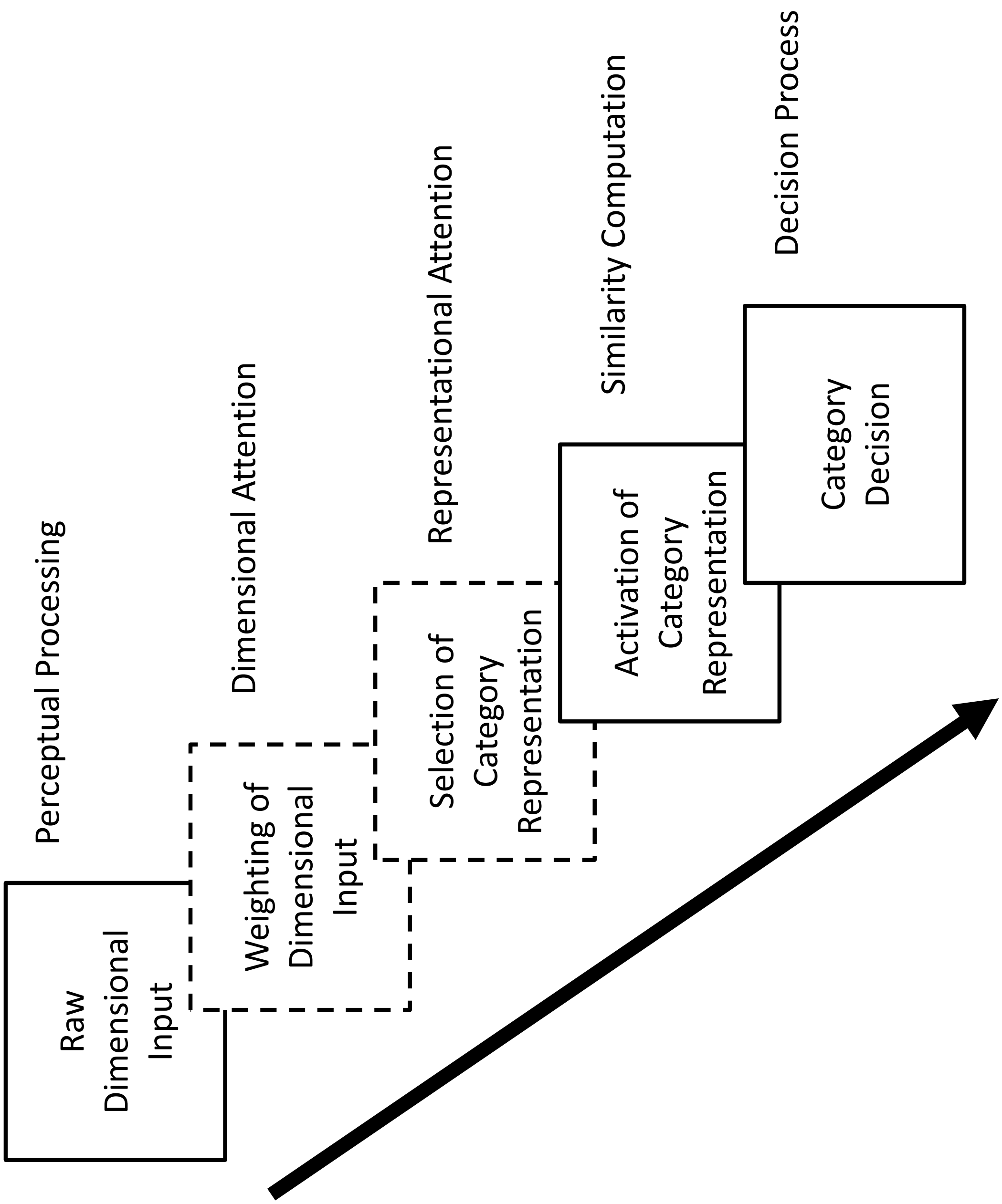
Figure 5. Category space used in Experiment 2. The abscissa describes the position of a vertically oriented bar along the base of the stimulus (relative to the center of the stimulus). The ordinate denotes rectangle height. Filled diamonds denote training stimuli, open squares denote transfer stimuli. Solid lines are the partial rule boundaries. Dashed lines divide the space into four diagnostic quadrants, which are numbered in the figure. Two example stimuli are shown underneath the category space.

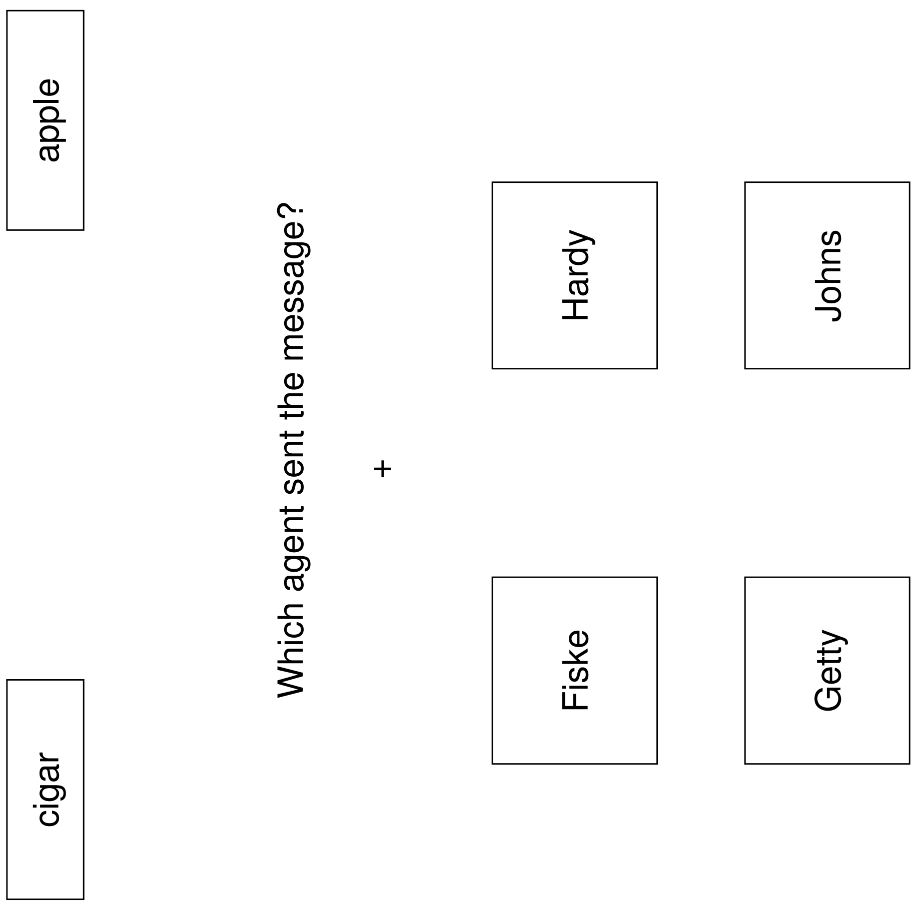
Figure 6. Ideal response profiles associated with the context-insensitive (CI; top row) and knowledge partitioning (KP; bottom row) strategies. Performance in the left and right contexts are shown in the left and right columns of panels, respectively.

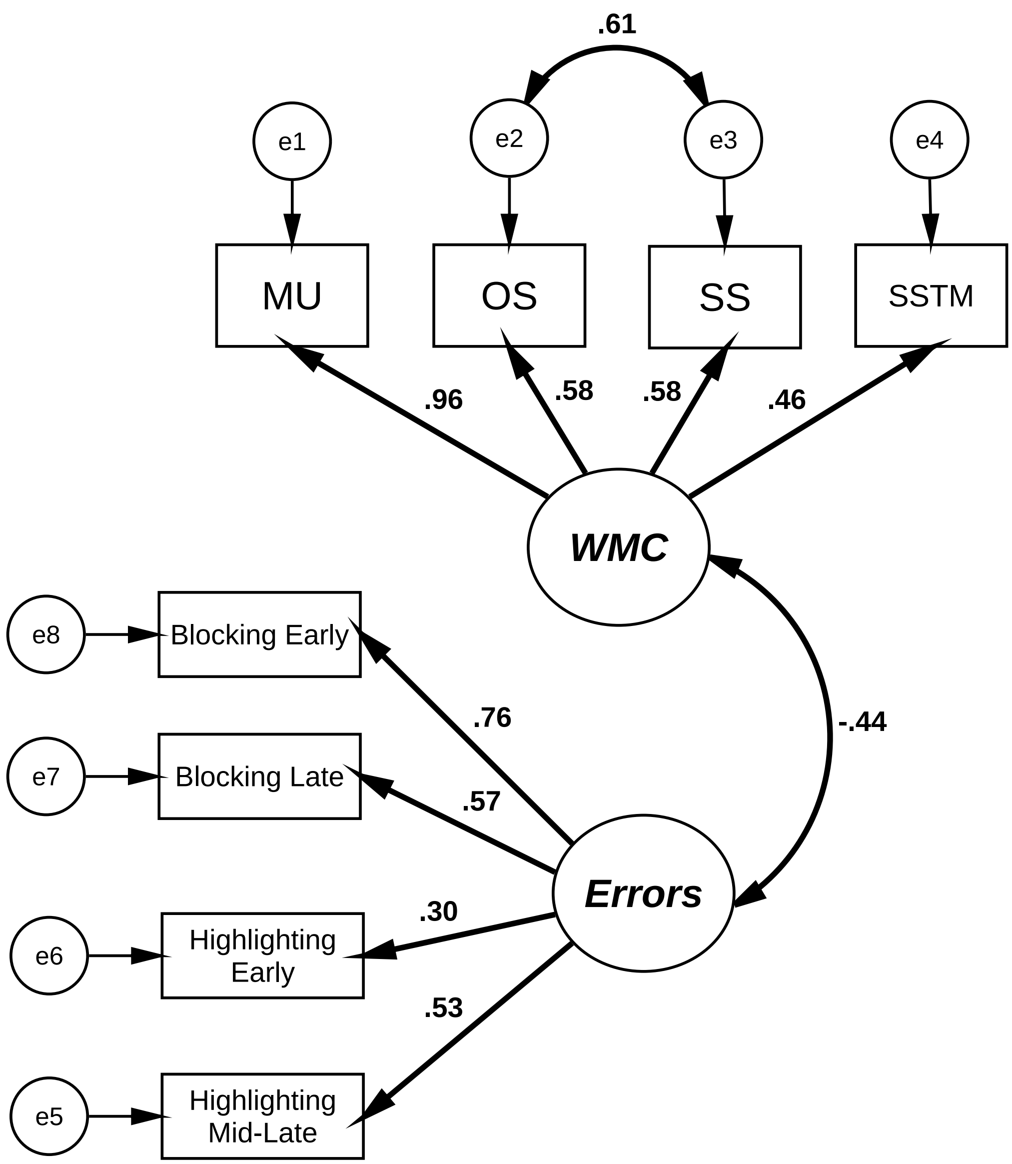
Figure 7. Item-wise P(A) in each context in the first transfer test for the KP-first and CI-first conditions in Experiment 2. Darker levels of shading correspond to higher P(A). Shading varies in steps of .1.

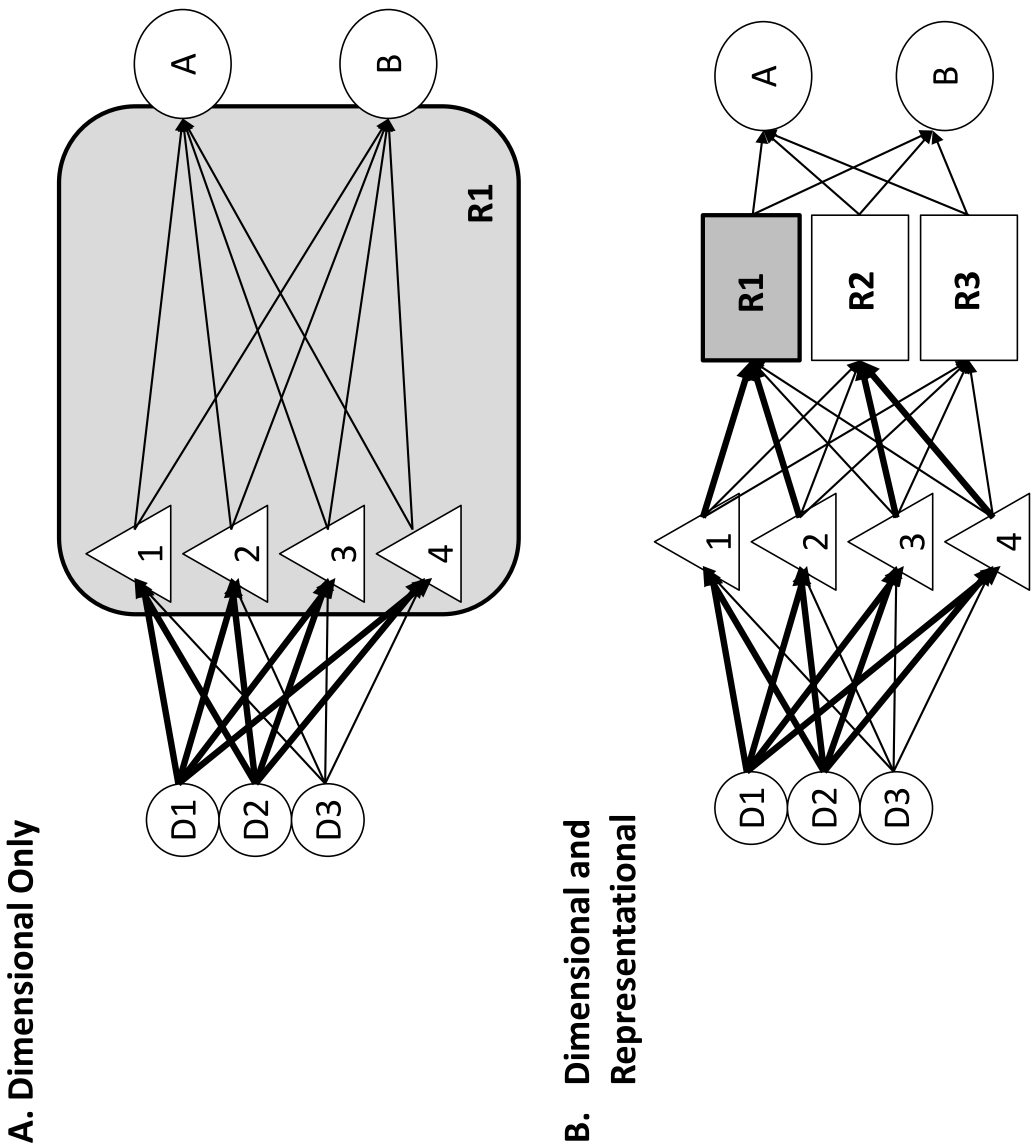
Figure 8. Context sensitivity across all transfer tests in Experiment 2 for the KP-first and CI-first conditions.

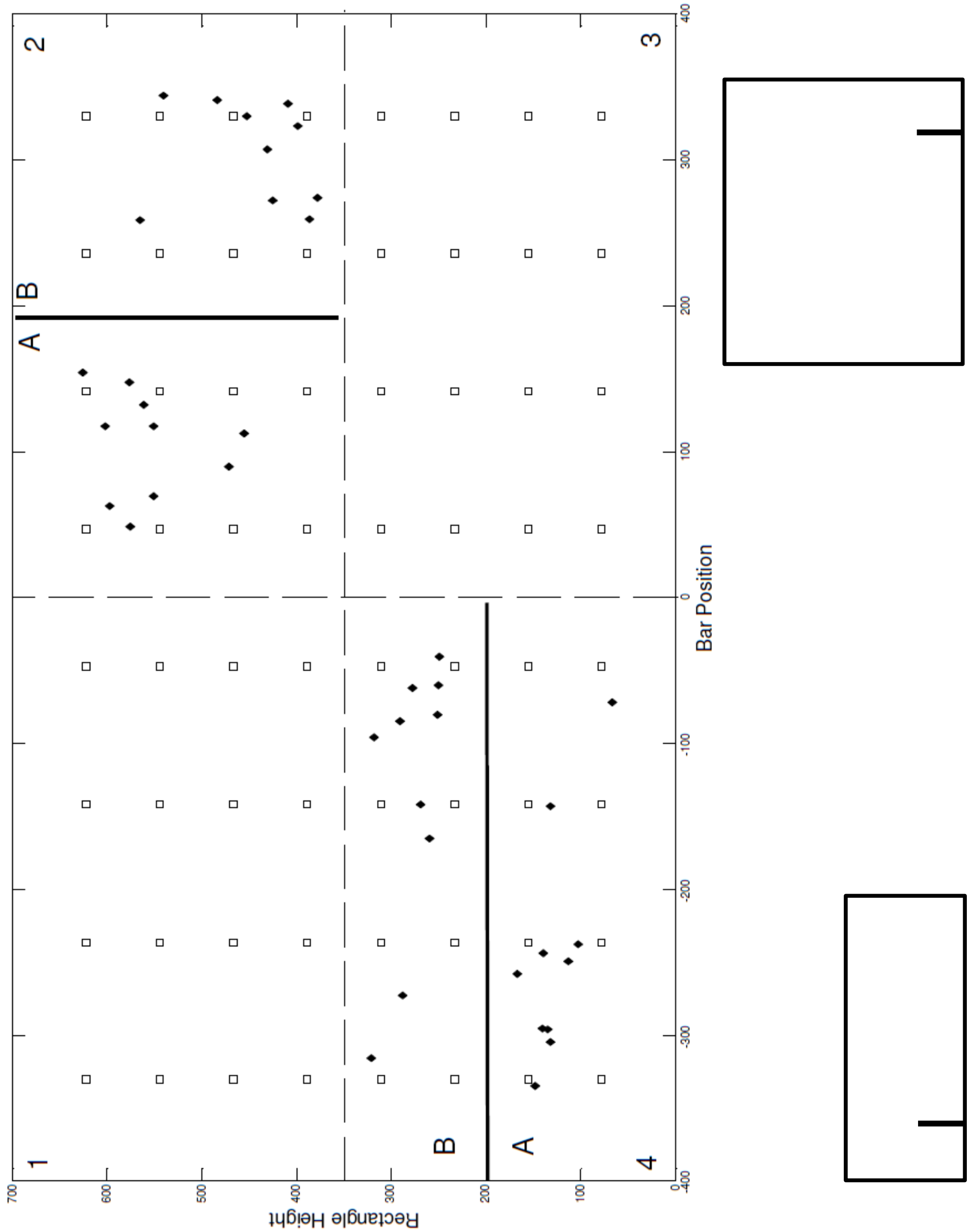
Figure 9. Structural model for Experiment 2. Significant correlations and factor loadings (all standardized estimates) are presented in bold.

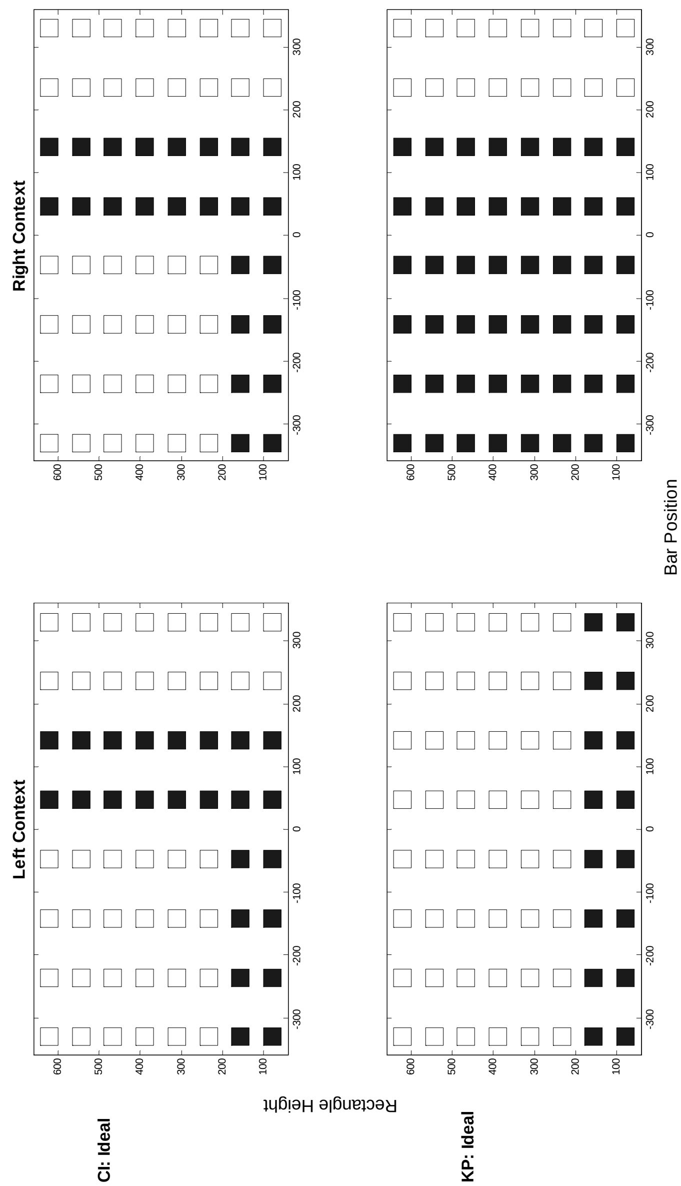

Raw Dimensional Input
Perceptual Processing
Weighting of Dimensional Input
Dimensional Attention
Selection of Category Representation
Representational Attention
Activation of Category Representation
Similarity Computation
Category Decision
Decision Process

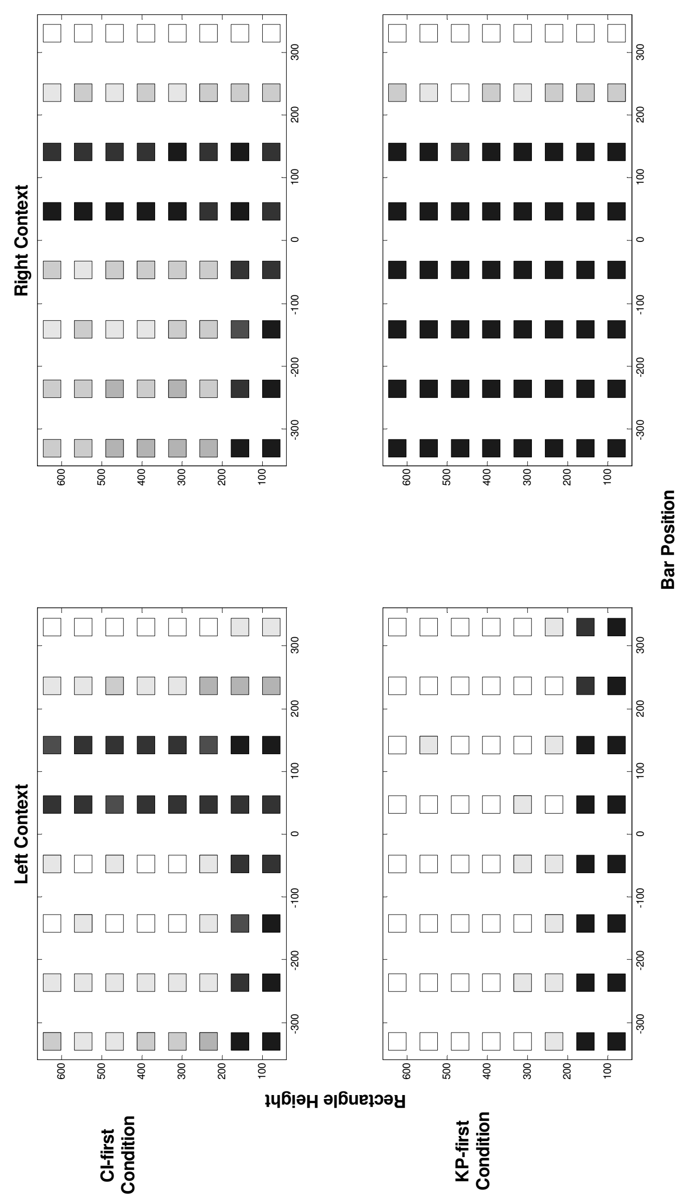

apple
cigar
Which agent sent the message?
+
Hardy
Fiske
Johns
Getty

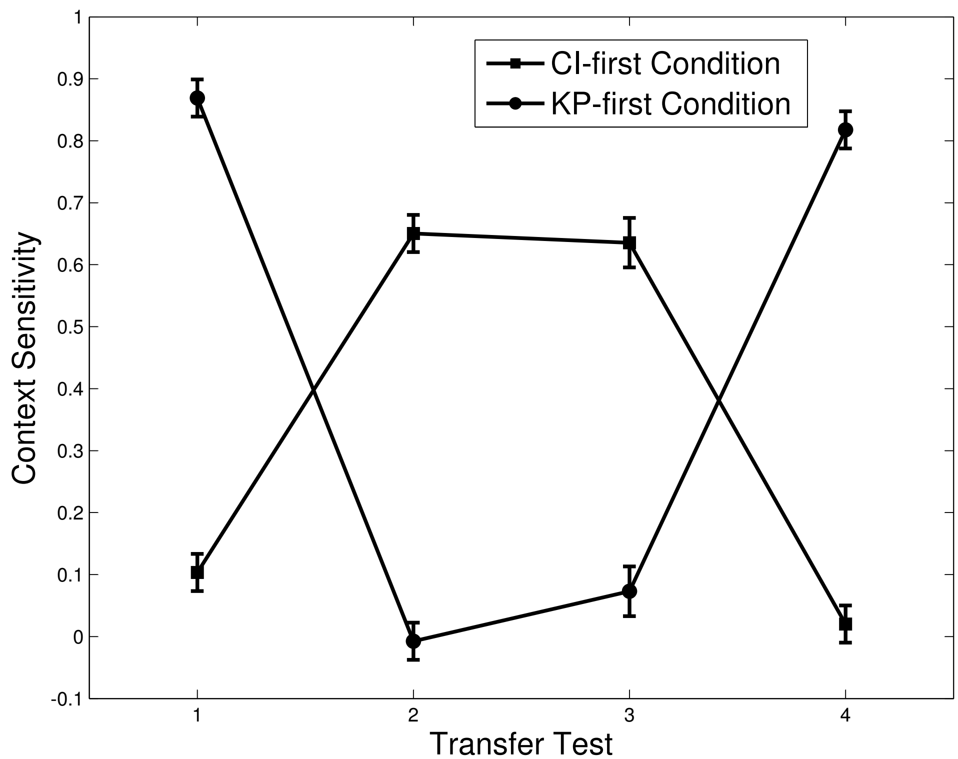

.61
e1
e2
e3
e4
MU
OS
SS
SSTM
.96
.58
.58
.46
WMC
-.44
e8
Blocking Early
e7
Blocking Late
.76
.57
Errors
.30
e6
Highlighting
Early
.53
e5
Highlighting
Mid-Late

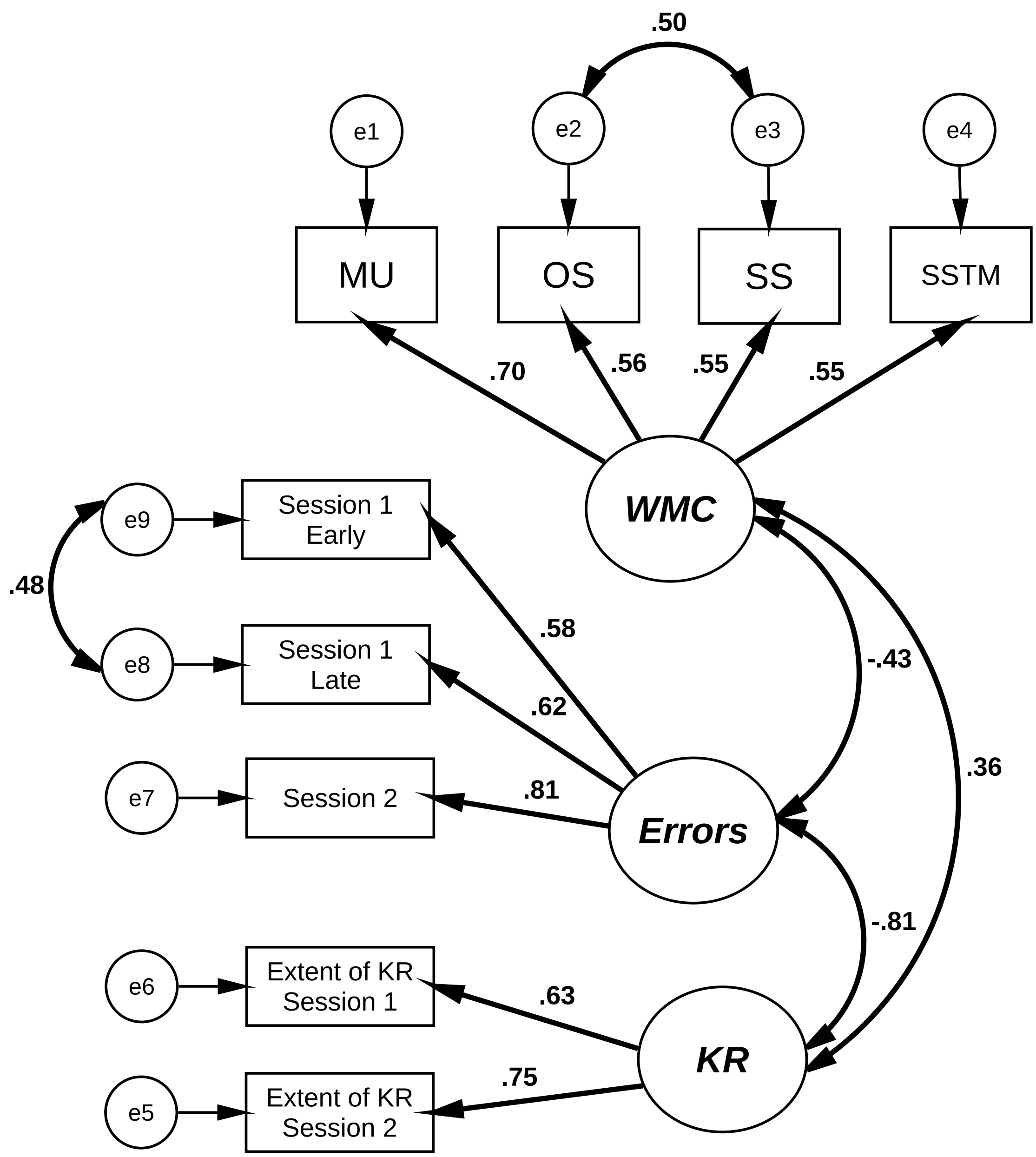

A. Dimensional Only
D1
D2
D3
1
2
3
4
A
B
R1
B. Dimensional and Representational
D1
D2
D3
1
2
3
4
R1
R2
R3
A
B


Rectangle Height
700
600
500
400
300
200
100
0
Bar Position
-400
-300
-200
-100
0
100
200
300
400
1
2
3
4
A
B
A
B


Left Context
Right Context
CI: Ideal
KP: Ideal
Rectangle Height
Bar Position
-300
-200
-100
0
100
200
300
100
200
300
400
500
600


Left Context
Right Context
CI-first Condition
KP-first Condition
Rectangle Height
Bar Position
600
500
400
300
200
100
-300
-200
-100
0
100
200
300


CI-first Condition
KP-first Condition
Context Sensitivity
Transfer Test
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
1
2
3
4


.50
e1
e2
e3
e4
MU
OS
SS
SSTM
.70
.56
.55
.55
WMC
e9
Session 1
Early
.48
e8
Session 1
Late
.58
-.43
.62
e7
Session 2
.81
.36
Errors
-.81
e6
Extent of KR
Session 1
.63
KR
e5
Extent of KR
Session 2
.75