Journal of Experimental Psychology:
Learning, Memory, and Cognition
2012, Vol. 38, No. 4, 881–904

0278-7393/12/$12.00 DOI: 10.1037/a0027298

Working Memory Does Not Dissociate Between Different Perceptual Categorization Tasks

Stephan Lewandowsky
University of Western Australia

Lee-Xieng Yang
National Chengchi University

Ben R. Newell
University of New South Wales

Michael L. Kalish
University of Louisiana at Lafayette

Working memory is crucial for many higher level cognitive functions, ranging from mental arithmetic to reasoning and problem solving. Likewise, the ability to learn and categorize novel concepts forms an indispensable part of human cognition. However, very little is known about the relationship between working memory and categorization. This article reports 2 studies that related people's working memory capacity (WMC) to their learning performance on multiple rule-based and information-integration perceptual categorization tasks. In both studies, structural equation modeling revealed a strong relationship between WMC and category learning irrespective of the requirement to integrate information across multiple perceptual dimensions. WMC was also uniformly related to people's ability to focus on the most task-appropriate strategy, regardless of whether or not that strategy involved information integration. Contrary to the predictions of the multiple systems view of categorization, working memory thus appears to underpin performance in both major classes of perceptual category-learning tasks.

Keywords: working memory, category learning, multiple memory systems, perceptual categorization

The ability to discriminate different objects and group them together in classes of similar entities—to categorize—is recognized as a fundamental aspect of human cognition (e.g., Ashby, Paul, & Maddox, 2011; Estes, 1994). Likewise, the ability to hold and manipulate information in memory for brief periods of time—to use working memory—is also considered an elemental facet of human functioning (Oberauer, 2009). Despite the status of these abilities as two pillars of cognition, until very recently, little research had focused on the theoretical and empirical relationship between the two. This article contributes to recent work that attempts to redress the balance (e.g., DeCaro, Thomas, & Beilock, 2008; Erickson, 2008; Lewandowsky, 2011).

There are several strong theoretical reasons for focusing on the relationship between working memory and category learning. First, exploring their empirical relationship may help adjudicate between competing theories of working memory—for example, whether it is best understood as executive attention (e.g., Kane, Bleckley, Conway, & Engle, 2001) or the ability to bind together temporary and transient representations (Oberauer, 2009). Any examination of category learning in the context of working memory thus ought to be of interest to working memory theoreticians more broadly; we discuss those broad implications after presentation of our data. Second, and more important in the context of this special section of the *Journal of Experimental Psychology: Learning, Memory, and Cognition,* examination of the relationship between working memory and categorization can shed light on the crucial debate over single- versus multiple-memory-system (MMS) views of category learning. Theoreticians who ascribe to the MMS view generally propose one explicit system that is reliant on working memory (specifically for the storage and testing of verbal rules) and one implicit system that does not require working memory but learns associations between (motor) responses and category labels (e.g., Ashby & Maddox, 2005; Maddox, Love, Glass, & Filoteo, 2008). The product of learning from the latter system is often assumed to be unavailable to awareness or impossible to verbalize (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Knowlton, Squire, & Gluck, 1994; Minda & Miles, 2010), and "working memory is not required . . . because the response is automatically linked with the feedback" (Filoteo, Lauritzen, & Maddox, 2010, p. 415).

Stephan Lewandowsky, School of Psychology, University of Western Australia, Perth, Australia; Lee-Xieng Yang, Department of Psychology and Research Center for Mind, Brain, and Learning, National Chengchi University, Taipei City, Taiwan; Ben R. Newell, School of Psychology, University of New South Wales, Sydney, Australia; Michael L. Kalish, Institute of Cognitive Science, University of Louisiana at Lafayette.

Preparation of this article was facilitated by a Discovery Grant from the Australian Research Council to the first author and Gilles Gignac, an Australian Professorial Fellowship to the first author, a Discovery Project Grant from the Australian Research Council to the third and fourth authors (and John Dunn), and a National Science Council, Taiwan, grant to the second author. We thank Chung-Yu Wang, I-Chia Chen, Sung-En Chien, Tzu-Yi Chung, Wan-Jung Lo, Chen-Yen Chung, and Weinn Jheng for assistance during data collection.

Correspondence concerning this article should be addressed to Stephan Lewandowsky, School of Psychology, University of Western Australia, Crawley, W.A. 6009, Australia. E-mail: stephan.lewandowsky@uwa.edu.au

Working Memory and Multiple Systems of Categorization

It is helpful to present a more complete picture of the presumed selective involvement of working memory, which is central to the MMS view and which is at the heart of this article. In principle, the differentiation between different memory systems is inextricably linked to the existence of different types of tasks, each of which is putatively linked to one, and only one, of the systems. For example, Ashby and O'Brien (2005) differentiated between four different tasks, each of which was presumed selectively to tap a different memory system. Here, we focus on two pairings between systems and tasks that have been particularly prominent in the literature: On the one hand, the explicit system, which requires the resources of working memory (Zeithamova & Maddox, 2006), is primarily responsible for the learning of tasks in which stimuli can be classified on the basis of simple verbalizable rules. Those tasks are known as rule-based tasks (RB tasks). On the other hand, the implicit system primarily supports performance in tasks for which verbal rules cannot—so it is claimed—be readily induced and applied (Minda & Miles, 2010). Those tasks are commonly referred to as information-integration tasks (II tasks).

From the inception of the MMS view, the presumed selective involvement of working memory in rule-based but not information-integration learning has been a core prediction of the theory. Maddox and Ashby (2004) reviewed the evidence for COVIS (competition between verbal and implicit systems), one of the leading MMS models, and described six a priori predictions all premised on the selective involvement hypothesis. For example, interpolating a delay between making a response and receiving feedback is claimed to impact II tasks more than RB tasks because, in the latter, working memory can be relied on to maintain and rehearse rules during the delay, whereas II tasks require almost immediate feedback to ensure reinforcement learning (Maddox & Ing, 2005). Similarly, II tasks are affected substantially by changes in the motor requirements for responding, but RB tasks are not, because performance is mediated by stimulus-specific rules held in working memory that are not tied to particular motor-response patterns (Ashby, Ell, & Waldron, 2003). The reliance on working memory is not always beneficial: A further prediction is that when time to process feedback is reduced (by placing an additional task immediately after the category feedback is presented), RB tasks are affected more than II tasks. The claim, once again, is that because only RB tasks rely on working memory resources, the reduction in time to process feedback impacts RB tasks but not the automatically learned II tasks (e.g., Filoteo et al., 2010). As Maddox et al. (2008) stated succinctly, all of these dissociation-based studies are used as evidence that "the procedural system does not interact with WM processes" (p. 580). Claims for the selective involvement of working memory in RB tasks extend to the neural level. For example, Maddox and Ashby (2004) stated, "The anterior cingulate selects new explicit rules to load into working memory" (p. 312).

Whether these dissociation-based studies and their accompanying neural specifications actually provide support for separable systems underlying information-integration and rule-based learning is keenly debated (Newell & Dunn, 2008; Newell, Dunn, & Kalish, 2011); what is beyond dispute, however, is that the selective involvement of working memory in rule-based but not information-integration learning has been crucial to the development of the MMS perspective, particularly the COVIS model from its origins (Ashby et al., 1998) to current speculation about the interactions of its proposed systems (Ashby & Maddox, 2011).

The preceding theoretical analysis clarifies the central distinguishing role played by working memory within the MMS view and, by implication, highlights the importance of determining the exact contribution of working memory to category learning.

Working Memory and Categorization: Extant Data

Despite this presumed dissociative role of working memory, much of the relevant supporting evidence to date comes from indirect measures, such as examining the effects of concurrent cognitive tasks during category learning (e.g., Foerde, Poldrack, & Knowlton, 2007; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). Briefly, those studies typically attempt to show that rule-based performance, but not information-integration performance, is impaired by a secondary task that ostensibly occupies working memory during training. For example, in the study by Zeithamova and Maddox (2006), participants in one condition had to perform an attention-demanding secondary task on each category-learning trial in addition to classifying the stimulus. On each trial, the stimulus was flanked by two digits, which differed both in value and in size. Following each categorization response, participants were asked to identify the side on which the larger digit was presented, with numeric and spatial magnitude queried unpredictably. Participants in the control condition did not perform this interfering task. Zeithamova and Maddox found that mean performance for the II and RB tasks differed much more for the secondary-task group than for the control group, suggesting that occupying working memory had a selective detrimental effect on RB performance.

However, evidence from those studies must be regarded with some caution for two reasons. First, in each of these cases, the original MMS interpretation of the reported results has been questioned by subsequent research (see, e.g., Lewandowsky, 2011; Newell, Dunn, & Kalish, 2010; Newell, Lagnado, & Shanks, 2007; Nosofsky & Kruschke, 2002; Nosofsky, Stanton, & Zaki, 2005). With respect to Zeithamova and Maddox's (2006) result, for example, Newell et al. (2010) showed that the differential influence of working memory load was due only to the inclusion in the analysis of participants who responded randomly. Such participants are typically and properly excluded from comparisons between tasks when the question at hand is how a task is learned. Newell et al. (2010) conducted a replication and two extensions of Zeithamova and Maddox's work and consistently demonstrated that participants did not show any differential influence of working memory load on RB and II tasks.

Second, concurrent-task studies suffer from the in-principle drawback that tradeoffs between the two simultaneous tasks are difficult to control or prevent (cf. Conway et al., 2005). It follows that differences in categorization performance may reflect uncontrolled tradeoffs between categorization and the secondary task. For example, Zeithamova and Maddox (2006) observed equal accuracy levels for the secondary tasks between rule-based and information-integration learning without, however, reporting the associated differences in latencies. Their observed differences in category-learning performance may therefore have been correlated with uncontrolled differences in secondary-task performance. Moreover, secondary tasks tend to engage only one facet of working memory (e.g., verbal or spatial working memory, but not both). Given the considerable evidence that working memory is a multifaceted theoretical construct (e.g., Oberauer, Süβ, Wilhelm, & Wittmann, 2003), a preferable approach to the problem might involve measurement of working memory capacity (WMC) and exploiting naturally occurring variation in WMC to extract the linkages between working memory and the two types of category-learning tasks.

A few studies exist that have included direct measurement of both WMC and category-learning ability. Arguably, however, those studies are subject to their own set of problems, and they have painted an equivocal picture of the differential involvement of working memory in RB and II tasks. For example, DeCaro et al. (2008) reported a dissociation between WMC and categorization performance, such that WMC was positively associated with performance on an RB task but negatively associated with performance on an II task. This outcome, of course, is highly supportive of the multiple-systems view, but it is also somewhat counterintuitive because it means that a construct that is known to be highly correlated with general ability (e.g., Kane, Hambrick, & Conway, 2005) can sometimes act to suppress learning performance. The results of DeCaro et al. have since undergone considerable reevaluation: Tharp and Pickering (2009) suggested that the outcome may have reflected an inappropriate performance measure, namely, an insufficient number of trials in a trials-to-criterion measure. Tharp and Pickering showed that this criterion was unacceptably lax and, thus, might have spuriously created a negative correlation between WMC and II task performance. In a small-scale replication of their earlier study, DeCaro, Carlson, Thomas, and Beilock (2009) confirmed that WMC was positively associated with performance in both RB and II tasks when a more suitable performance criterion was used. On balance, therefore, the work of DeCaro and colleagues provides a contradictory picture.

More recently, Lewandowsky (2011) related performance on all six problems from the seminal study of Shepard, Hovland, and Jenkins (1961) to WMC and found through structural equation modeling that a single latent variable was sufficient to capture individual performance differences for all problem types. Moreover, this latent variable was highly correlated with another latent variable representing WMC as measured by a battery of four tasks that spanned the verbal and spatial domains. Given that some of the problem types in Shepard et al. map onto RB tasks, whereas others translate into II tasks (Minda, Desroches, & Church, 2008), the uniform association between WMC and learning performance reported by Lewandowsky can be seen as a challenge to the multiple-systems view. However, as in the studies by DeCaro and colleagues (DeCaro et al., 2008, 2009), Lewandowsky used stimuli comprised of discrete binary dimensions, which may have limited the extent to which his data can be interpreted with reference to the MMS view for two reasons. First, it is known that when an experiment uses few (eight, in this case) stimuli that are composed of binary dimensions, people may solve the tasks based on memorization of exemplars (Rouder & Ratcliff, 2004). Memorization may thus obscure rule use or information integration. Second, with discrete and discriminable stimuli, partial rules can support above-chance performance even with multidimensional category spaces that purport to require information integration (Lewandowsky, Roberts, & Yang, 2006; Tharp & Pickering, 2009). Stimuli of that type thus further diffuse the demarcation between II and RB tasks because they may facilitate a (partial) rule-based solution even involving II tasks.

In summary, it remains uncertain whether working memory is selectively involved in rule-based categorization or whether it is generally involved in all forms of learning. This article reports two experiments that test the prediction of the MMS view that working memory should mediate rule-based performance but not information-integration learning.

Testing the MMS View

The present methodology was designed to resolve several of the problems and limitations associated with existing research. First, we used stimuli that, unlike those used in all previous studies involving WMC, are known to largely defy memorization of exemplars (Rouder & Ratcliff, 2004). We used highly perceptual stimuli that have continuous dimensions, namely, Gabor patches. If memorization of exemplars is difficult or impossible, potential differences between rule use and information integration are more readily observable because they are less likely to be obscured by an exemplar-driven process.

Second, category spaces created by Gabor patches are said to prevent or at least impair verbal characterization (e.g., Markman, Maddox, & Worthy, 2006; Zeithamova & Maddox, 2007). This is particularly important in the present context because related precedents are arguably permitted at least partial verbalization of rules, even in II tasks (e.g., DeCaro et al., 2008; Lewandowsky, 2011).

The possibility of partial verbalization leads us to our third critical feature of the methodology. This feature requires clarification of our nomenclature. Thus far, we have referred to the two types of *tasks* as RB (rule-based) and II (information-integration) tasks, respectively, and we continue to use that terminology. However, from here on, we additionally characterize people's imputed cognitive *strategies* by using the terms *rule use* and *information integration,* respectively. This distinction is necessary because although RB and II tasks typically elicit rule use and information integration, respectively, this mapping cannot always be taken for granted. In particular, it is not uncommon for people to solve II tasks by resorting to the application of simple rules. To illustrate, Maddox et al. (2008) trained participants to label four categories of lines that varied in length and orientation. When members of one group of participants guessed incorrectly during training, they received only the information that they were incorrect (negative reinforcement, or partial feedback). When a member of the other group (full feedback) guessed incorrectly, they were given the correct category label along with the negative feedback. Maddox et al. showed that providing full feedback decreased performance on II tasks, a result they ascribed to people's tendency to fall back onto simple rules—rather than relying on information integration—when given full feedback. In confirmation, Maddox et al. fit verbalizable strategies to each participant's data and found a 50% increase in the number of participants whose information-integration performance was best described by rule use when feedback was partial (39%) instead of full (26%).

If people apply rules when solving an II task, then any comparison between the two tasks would not properly compare the two purported memory systems. To properly test the MMS view, it is therefore crucial to differentiate between the physical task parameters (RB vs. II) on the one hand and the psychological process (rule use vs. information integration) that people engage to solve a task on the other. In our experiments, we characterized each person's responses by fitting competing models that embodied either rule use or information integration (or indeed several other possible strategies). We were therefore able to draw comparisons between rule use and information integration at a psychological level, without having to be concerned with possible contamination of II task performance by partial verbal rules. A further beneficial consequence of the identification of individual strategies is that proponents of the MMS view would find it more difficult to reject possibly challenging findings—namely, a lack of dissociation between RB and II task performance—by appealing to partial rule use in the II task.

A fourth critical feature of our methodology arises from recognition of the fact that working memory spans multiple domains (Oberauer et al., 2003; Oberauer, Süβ, Schulze, Wilhelm, & Wittmann, 2000). Therefore, to measure WMC, we used a battery of several tasks that were explicitly designed to span the verbal and spatial domains (Lewandowsky, Oberauer, Yang, & Ecker, 2010). Likewise, we created multiple instantiations of each class of categorization task (RB and II), thereby permitting use of structural equation modeling (SEM) to examine the relationship between three potential latent variables, namely, WMC, RB task performance, and II task performance. SEM requires the presence of multiple indicator variables (e.g., multiple RB/II tasks performed by the same participants) to identify latent variables whose correlation with other constructs (e.g., WMC) can then be ascertained with little or no contamination by measurement error and task idiosyncracies (Coffman & MacCallum, 2005).

We chose to create multiple instantiations of the categorization tasks by manipulating their difficulty. Difficulty can be manipulated by placing exemplars within each category at different distances from the corresponding category boundary (Figure 1 shows the specific category spaces used in our first study). The distance to boundary permits fine-tuning of the ease with which exemplars from the different categories can be discriminated, thus manipulating overall task performance within each class of tasks. An additional beneficial side effect of using difficulty to create multiple instantiations of II and RB tasks is that it can (at least roughly) equalize average performance between the two classes of tasks. This equalization guards against the possibility that the relationship between WMC and categorization performance might differ between regions of the difficulty space, which could artifactually introduce a dissociation where there is none.

To foreshadow our results, in both experiments, learning performance was found to be related to WMC, in both II and RB tasks. Moreover, a competing SEM that instantiated the predictions of the multiple-systems view—namely, that WMC should be uncorrelated with II task performance, whereas WMC should be associated with RB task performance—was decidedly rejected in

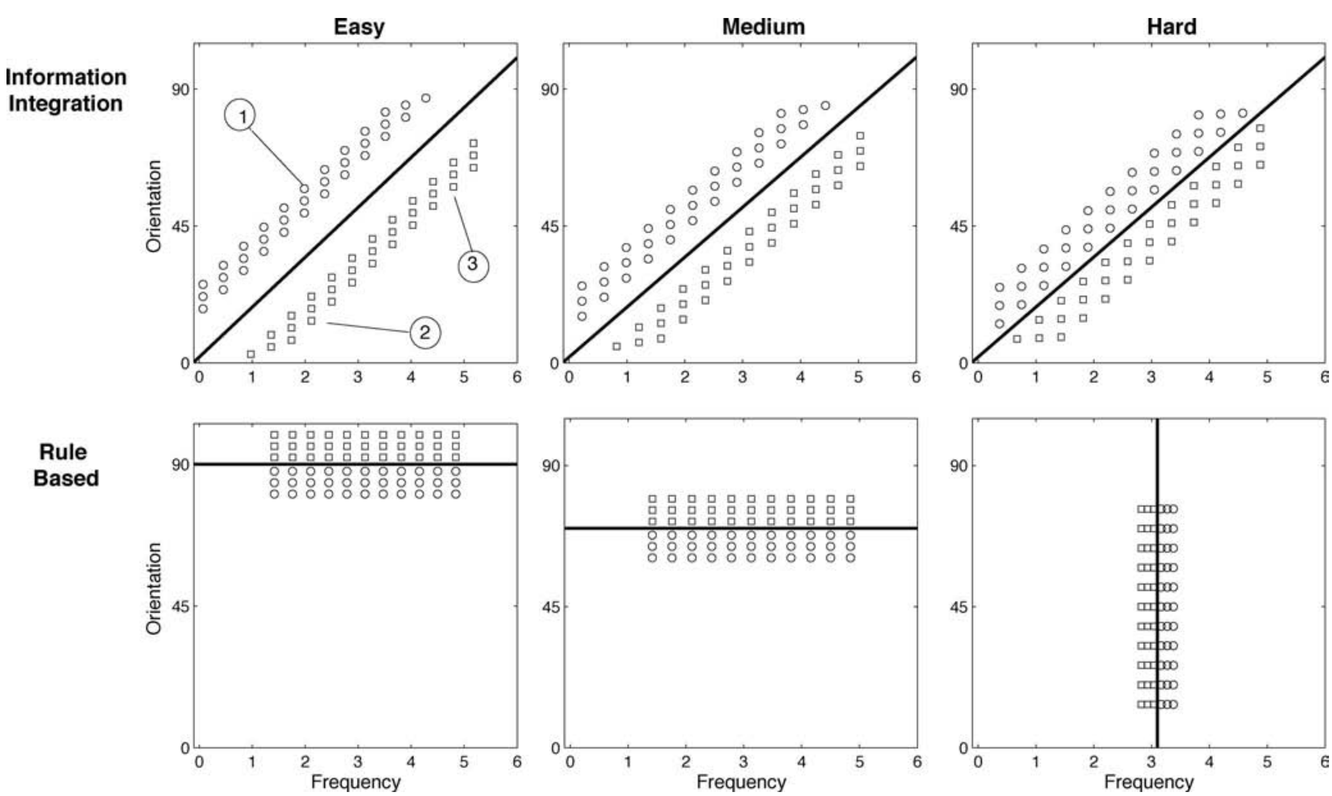

Figure 1. The category structures for all six tasks used in Experiment 1, with information-integration tasks in the top row of panels and rule-based tasks in the bottom row. Columns refer to, from left to right, easy, medium, and hard problems. The three stimuli that are numbered within callouts in the top left panel are shown in Figure 2.

both studies. The positive mediating role of WMC extended to an analysis in which each individual's performance was expressed as the likelihood of having used a rule-based or information-integration strategy: Although the preferred choice of strategy differed between tasks—with one-dimensional rule use predominating in RB tasks and information-integration strategies being more prevalent in II tasks—increasing WMC facilitated use of both strategies. The data provide no evidence for the suggestion that working memory is selectively involved in some types of perceptual category learning but not others.

Experiment 1

Method

Participants and apparatus. One hundred and eleven undergraduate students at National Chengchi University participated in this experiment. The participants were paid NT$150 (≅ US$5) for each session (about 1.5 hr). Every participant completed the three sessions on 3 different days within 2 weeks. Additionally, to encourage use of the optimal categorization boundary, a bonus of NT$50 (≅ US$1.70) was paid if accuracy in any one of the learning blocks exceeded a criterion (explained later). The bonus was not paid more than once for each category-learning task. Across the six tasks (2 types $\times$ 3 difficulties), the maximum achievable bonus was thus NT$750 (≅ US$25).

The experiments in this article were controlled by a Windows PC running a Matlab program designed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Categorization stimuli. The category spaces for all six tasks are shown in Figure 1, and three sample stimuli are shown in Figure 2. The locations of the three sample stimuli in the category space are indicated by numbers in callouts in the top-left panel of Figure 1. Stimuli varied along two dimensions, namely, orientation and spatial frequency of the grating in the Gabor patches.

The II tasks were characterized by the fact that both dimensions had to be considered in a decision, whereas for the rule-based tasks a single dimension was sufficient. Difficulty was manipulated by changing the separation between the categories along the diagonal boundary for the II tasks and by changing the location of the boundary for RB tasks. Zeithamova and Maddox (2007) showed that a boundary along the orientation dimension at 70° was more difficult to learn than one at 90° (i.e., along a cardinal axis) and that a boundary using spatial frequency was more difficult still. Thus, for the easiest RB task, the two categories consisted of stimuli in which the orientation of the grating was either just clockwise or just counterclockwise of the horizontal (i.e., 90°); for the medium RB task, the two categories likewise straddled a boundary at an orientation of 70°; and for the hardest RB task, orientation was irrelevant, and the two categories were defined by differences in spatial frequency only. Pilot testing confirmed the order of difficulty among all three levels of difficulty of each task. From here on, we identify tasks by combining their type (II vs. RB) with easy (EZ), medium (MED), or hard (HD) to indicate task difficulty (i.e., II-HD vs. RB-EZ, etc.).

Any diagonal boundary can, of course, be roughly approximated by conjunction or disjunction of two partial linear boundaries. For example, a person might form a rule such as, "if the frequency is greater than X and the orientation is less than Y, then the stimulus

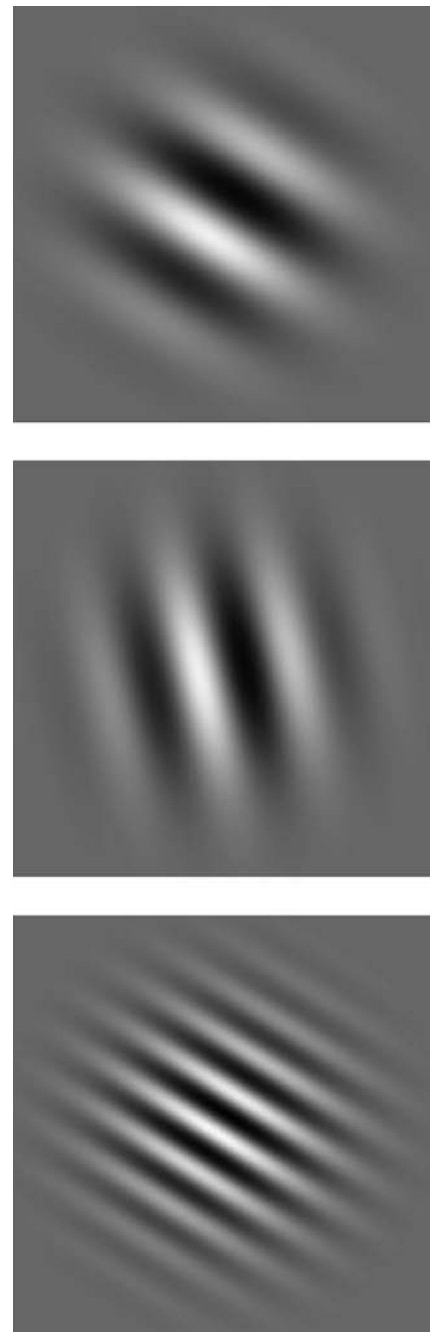

Figure 2. Three sample Gabor patches used as stimuli in the present experiments. From top to bottom, the stimuli correspond to those numbered 1, 2, and 3, respectively, in the top left panel of Figure 1.

is in Category A; otherwise it is in B." If X and Y are chosen optimally, bilinear rules can be quite accurate. We found the most accurate rule for each II task condition by grid searching all possible bilinear rules, which identified a maximum bilinear performance of .76 for the II-EZ task condition, .73 for the II-MED task condition, and .68 for the II-HD task condition. The bonus criteria during training were set equal to those maximum bilinear values for the II tasks. For the RB tasks, the bonus criterion was a uniform .90.

Procedure.

WMC battery. WMC was measured with the battery presented by Lewandowsky et al. (2010), except that the sentence-span (SS) task was omitted owing to time constraints. This left three tasks with which to measure WMC: an operation span (OS) task, a spatial short-term memory (SSTM) task, and a memory-updating (MU) task. Those tasks are described in detail in Lewandowsky et al. and are outlined briefly here.

The MU task required participants to (a) store a series of digits in memory, (b) mentally update those digits on the basis of a series of arithmetic operations, and (c) recall the updated digits. On each trial, three to five frames were presented that each contained a random digit. Successive arithmetic operations (e.g., $+4$ or -3) were then presented in the frames, one at a time, until people had to recall the contents of all frames after a varying number of steps. There were a total of 15 trials.

On each trial of the OS task, a series of arithmetic equations was presented (e.g., $4 + 3 = 7$), each of which was followed by a consonant. Participants judged the equation for correctness and memorized the consonants for later recall. A trial involved between four and eight unique consonants, which participants had to recall immediately after presentation in the original order. There were 15 trials total, with three trials for each set size.

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

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

Categorization tasks. Each categorization task comprised six learning blocks, each of which involved presentation of 66 stimuli—shown in the corresponding panel of Figure 1—in a random order. Each trial involved presentation of the appropriate Gabor patch in the center of the screen in a 200×200 pixel area. The stimulus remained visible for a maximum of 1,000 ms or until the participant responded. Responses were registered by pressing the *S* key or the semicolon key, respectively, for the two categories and were followed by corrective feedback (consisting of the word *correct* or *wrong*) for 500 ms. Trials were separated by a 2-s blank screen.

Scheduling of tasks and sessions. Tasks were assigned to the three experimental sessions in two possible sequences. In Sequence 1, the tasks were spread across sessions as follows: RB-EZ → MU → II-HD; II-MED → OS → RB-MED; and II-EZ → SSTM → RB-HD, for Sessions 1, 2, and 3, respectively. In Sequence 2, the assignment of tasks to sessions was RB-HD → MU → I-EZ; II-MED → OS → RB-MED; and RB-EZ → SSTM → II-HD, respectively. Participants were randomly assigned to sequence. In addition, the order of sessions was reversed for a random half of the participants within a sequence, thus yielding a total of four different task orders.

This sequencing method was intended to strike a balance between the need to maximize experimental control by complete counterbalancing and the desire to reduce method variance by minimizing variation in unnecessary experimental parameters across participants. This method follows precedent (Lewandowsky, 2011) and permits the removal of order effects through dummy regression, as we show later.

Results

Working memory capacity. Inspection of distributions identified two participants who were excluded from further analysis because at least one of their WMC scores was more than 3 standard deviations below the mean. Thus, the analyses relied on data from 109 participants. Summary statistics for the WMC tasks are shown in Table 1, and the correlations among them and with the categorization variables used in the structural equation model below are shown in Table 2. Table 1 also shows reliability estimates for the three WMC tasks, obtained by computing Cronbach's α across thirds of trials for each task.

Categorization. Average performance across blocks is shown in Figure 3 for all six tasks. In addition to the obvious effects of practice and task difficulty, the figure reveals that although there was considerable overlap between some tasks of each type, the RB tasks overall led to better performance ($M =$ .84) than the II tasks ($M = .71$). The figure shows that performance was significantly above chance even for the most difficult II-HD task (error bars are 95% confidence intervals and do not overlap with the dashed horizontal line representing chance), and it also shows that performance on the II-HD task, unlike for the other two II tasks, fell below the maximum value achievable with a bilinear boundary (isolated plotting symbols on the right of the figure).

In confirmation of these apparent effects, a 2 (task type: II vs. RB) $\times$ 3 (difficulty) $\times$ 6 (block) within-subjects analysis of variance (ANOVA) revealed significant main effects for task, $F(1, 108) = 410.36$, $MSE = 0.04$, $p < .01$; difficulty, $F(2, 216) = 328.07$, $MSE = 0.04$, $p < .01$; and block, $F(5, 540) = 403.30$, $MSE = 0.004$, $p < .01$. The two-way interactions involving Task $\times$ Difficulty, $F(2, 216) = 19.42$, $MSE = 0.04$, $p < .01$; Difficulty $\times$ Block, $F(10, 1080) = 2.42$, $MSE = 0.004$, $p < .01$; and Task $\times$ Block, $F(5, 540) = 2.84$, $MSE = 0.005$, $p < .05$, were also significant. Finally, the overarching three-way interaction was significant, $F(10, 1080) = 26.57$, $MSE = 0.004$, $p < .01$. Because focus here is on individual differences and the involvement of working memory, we do not consider those interactions further.

Table 1
Summary of Working Memory Capacity Scores in Experiment 1

Measure	MU	OS	SSTM
M	0.81	0.77	0.90
SD	0.16	0.13	0.05
Minimum	0.12	0.29	0.73
Maximum	1.00	0.98	0.98
Skewness	−1.52	−1.25	−0.67
Kurtosis	5.81	5.16	3.82
Cronbach's α	0.91	0.82	0.86
Standardized loadings	0.46	0.44	0.51

Note. Standardized loadings refer to working memory capacity measurement model. MU = memory-updating task; OS = operation span task; SSTM = spatial short-term memory task.

Table 2
Correlations Between WMC Tasks and Residualized Category Learning Measures Used in the Structural Equation Modeling for Experiment 1

Task	1	2	3	4	5	6	7	8	9	10	11
1. MU	—										
2. OS	.199*	—									
3. SSTM	.233*	.222*	—								
4. rRB-MEDe	.079	.153	.150	—							
5. rRB-MEDo	.134	.108	.137	.942**	—						
6. rRB-HDe	.121	.253**	.094	.318**	.265**	—					
7. rRB-HDo	.113	.293**	.065	.387**	.327**	.942**	—				
8. rII-EZe	.163	.161	.205*	.296**	.289**	.341**	.359**	—			
9. rII-EZo	.178	.194*	.145	.261**	.251**	.344**	.382**	.908**	—		
10. rII-MEDe	.166	.256**	.157	.415**	.353**	.299**	.334**	.572**	.594**	—	
11. rII-MEDo	.251**	.306**	.170	.400**	.347**	.346**	.378**	.680**	.681**	.898**	—

Note. MU = memory-updating task; OS = operation span task; SSTM = spatial short-term memory task; the prefix r = residualized; RB = rule-based task; II = information-integration task; MED = medium; HD = hard; EZ = easy; the suffix e = even-numbered trials; the suffix o = odd-numbered trials.
* $p < .05$. ** $p < .01$.

To examine the influence of task sequence on category learning, the difference between the two sequences was examined by a 2 (type of task) × 3 (difficulty) × 2 (sequence) between–within ANOVA. The analysis again showed the effects of type of task and difficulty already noted but additionally clarified that sequence did not determine performance: Neither the main effect of sequence nor any of the interactions to which it contributed were significant (all $Fs < 1$).

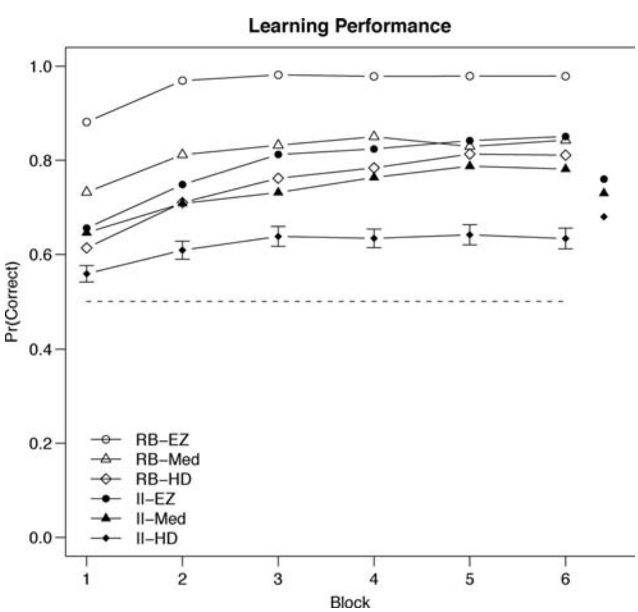

Figure 3. Proportion correct across blocks in Experiment 1 for all categorization tasks. Filled plotting symbols are for information-integration (II) tasks, and open plotting symbols are for rule-based (RB) tasks. The error bars for the II-HD task represent 95% confidence intervals (not shown for the other tasks to avoid clutter). The horizontal dashed line represents chance performance (50%) and the isolated filled plotting symbols on the right indicate the maximum performance achievable with a bilinear boundary for the II tasks. EZ = easy; Med = medium; HD = hard.

Structural equation modeling.

Brief overview. The basic idea of SEM is to create an efficient representation of the overall variance–covariance matrix relating all measured variables—in our case, the indicators of WMC and the performance measures for the categorization tasks. This process typically involves two stages: First, measurement models are created to capture the internal structure among a group of indicator variables that are conceptually linked, such as the measures of WMC. Measurement models contain at least one latent variable that captures the shared variance among indicators and that is resistant to measurement error and method variance that contaminates each indicator variable in isolation (Tomarken & Waller, 2005). The second stage involves creation of a structural model that relates the various latent variables created at the first stage, thereby identifying the relationships between the psychological constructs of interest. For the present study, we first created separate measurement models for WMC and category-learning performance, before combining those constructs into structural models that explicated the role of WMC in categorization.

In particular, we first constructed an unconstrained structural model in which the association between WMC and the two types of category-learning tasks, RB and II, was free to vary. We then compared that unconstrained model to two constrained nested models: one in which WMC and II task performance was forced to be uncorrelated, as expected by the MMS view, and another one in which both tasks were constrained to be equally correlated with WMC. Standard model comparison techniques can thereby provide a powerful test of the MMS view, because it expects the constrained model in which II task performance is uncorrelated with WMC to fit as well as the unconstrained model—a rejection of that model thus constitutes strong statistical evidence against the MMS prediction. This rejection of the prediction would be compounded if the model in which both correlations were constrained to be equal fit no worse than the unconstrained model.

A crucial aspect of SEM is to assess the model's goodness of fit, that is, the extent to which the final model suffices to recreate the entire variance–covariance matrix. Numerous measures of fit exist (MacDonald & Christiansen, 2002), and here we report four measures that are in common usage: (a) The model's chi-square statistic captures the deviation between the observed covariance matrix and the matrix imputed by the model. A significant chi-square suggests that the deviation is greater than would be expected by chance. (b) The comparative fit index (CFI) expresses model fit as a proportion of improvement relative to a model that assumes that all indicator variables are uncorrelated. The CFI ranges from 0 to 1, where 1 indicates maximal fit. Conventionally, CFI values greater than .90 (ideally, greater than .95) are considered to represent a good fit. (c) The root-mean-square error of approximation (RMSEA) represents the model's discrepancy per degree of freedom; it therefore arguably corrects for model complexity. In a well fitting model, RMSEA is .08 or less (ideally, less than .05), and the 90% confidence interval of the RMSEA should ideally range from 0 (or close to it) to no more than .10 (ideally, .08). (d) The final measure of fit, the standardized root-mean-squared residual (SRMR) refers to the average difference between standardized model-imputed and observed variances and covariances. Any value below .08 (ideally, below .05) is considered to reflect good fit. Interpretation of the various fit indices is subject to discretion and ongoing debate; our criteria are largely consonant with relevant precedent and discussions in the literature (e.g., Kline, 2005; McDonald & Ho, 2002; Wilhelm & Oberauer, 2006).

Measurement models. With three manifest variables (OS, MU, and SSTM), the measurement model for WMC was just identifiable with 0 *df.* The loadings of the three tasks on the WMC latent variable are shown in Table 1.

For the SEM analysis, the categorization data were preprocessed as follows. First, adopting relevant precedent (Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009; Wilhelm & Oberauer, 2006), each task was represented by two manifest variables: One of those manifest variables represented performance (i.e., proportion correct) on all odd trials (i.e., 1, 3, 5, . . .), and the other one represented performance on all even trials (2, 4, . . .). The use of multiple manifest variables rendered separate factors for RB and II tasks identifiable.

Second, those manifest variables were converted to residuals for the SEM analysis by removing the effects of counterbalancing sequence through dummy regression (cf. Wilhelm & Oberauer, 2006). Specifically, for each task, performance on all blocks was regressed onto four predictors that dummy coded (1 or 0) the sequence administered to each participant. The residuals of that analysis were added to the mean for that task across participants, thus yielding a set of transformed scores that (a) preserved all individual differences and mean overall performance while (b) removing practice effects and other method variance associated with the counterbalancing sequence. This residualized data set yields the same means as those shown in Figure 3 but simply removes the contribution of the counterbalancing scheme to individual differences in categorization performance. The correlations in Table 2 are based on the residualized data.

The measurement model we considered for the categorization tasks included two latent variables: one for the RB task and one for the II task. In addition, the pairwise correlations between the residuals of the manifest variables for each task (based on odd and even trials) were freely estimated. This model fit very well, $\chi^2(47) = 53.88$, $p > .1$, CFI = .995, RMSEA = .037 (90% CI [.0, .076]), SRMR = .0679. The correlation between the latent variables for II and RB was high and significant, $r = .68$, $Z = 7.29$, $p < .0001$.

Structural model. Both measurement models were combined into a final unconstrained structural model involving three latent variables (WMC, II, and RB). This model fit extremely well, $\chi^2(81) = 86.92$, $p > .1$, CFI = .996, RMSEA = .026 (90% CI [.0, .061]), SRMR = .0639, and provided the basis against which to examine the MMS hypothesis. To test the hypothesis, we first compared the unconstrained model with one in which WMC was uncorrelated with II task performance by setting the correlation between the corresponding latent variables to zero. This model fit significantly worse, $\Delta\chi^2(1) = 13.426$, $p < .0005$. We next compared the unconstrained model with one in which both correlations involving the WMC latent variable were constrained to be equal. This model's fit did not differ from that of the unconstrained model, $\Delta\chi^2(1) = 0.212$, $p > .10$, and its fit was also excellent by absolute standards, $\chi^2(82) = 87.132$, $p > .1$, CFI = .996, RMSEA = .024 (90% CI [.0, .060]), SRMR = .0632. The correlation between WMC and the two latent variables representing the categorization tasks was substantial, $r = .518$, and significant, $Z = 2.551$, $p \approx .01$. We therefore adopted the model with equal correlations between WMC and both types of categorization tasks as our final structural model (see Figure 4).

Response modeling. In an effort to characterize the categorization strategies that people were using during the experiment, we fit a set of response surface models to the data from each individual. Because the response surface models were fit to the raw binary responses (i.e., sequences of A or B responses, represented as 0s or 1s), the data could not be residualized for this analysis.

For each participant, each training block was fit by six different models representing strategies that people are known to be able to apply to problems of this type (see the Appendix for details). The two simplest models were one-dimensional classifiers, with the boundary between the two categories instantiated by a line parallel to one of the two dimensional axes. Each of these one-dimensional rule-based models had two parameters: the location of the boundary and the rate at which response probabilities changed across it (which could reflect either perceptual or criterial noise).

We also fit a general linear boundary, which has two location parameters and a noise term. This model corresponds most closely to the presumed information integration because both dimensions are considered jointly and in an integrative manner to characterize people's responding.

We further fit two bilinear models: one instantiating a conjunctive rule and another one that embodied a disjunctive rule. Both models involved two partial boundaries at right angles to one dimension, but they differed with respect to how the two partial boundaries were integrated; see Appendix for details. These models required location parameters for each of the two one-dimensional component boundaries and (to equate complexity with the general linear bound) a common error term.

Finally, we fit a random-response model, which used a single parameter representing the probability of labeling an item as A but did not in any way differentiate between the two categories.

All models were optimized using log likelihood, with multiple starting values. The likelihoods were converted to penalized likelihood functions (Bayesian information criterion [BIC]; Schwarz, 1978) to allow direct comparison of models with different numbers

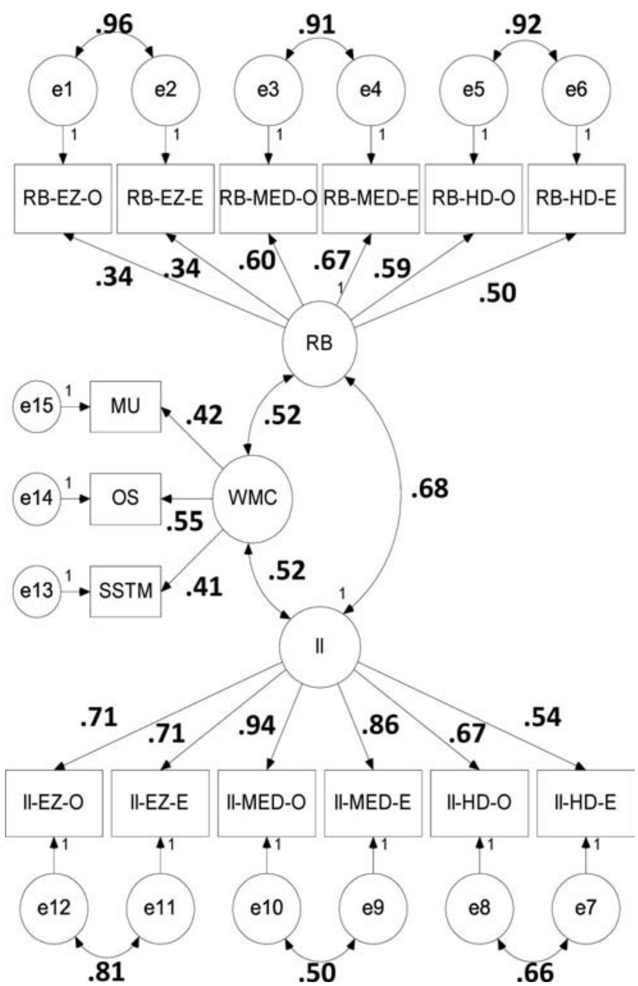

Figure 4. Structural model relating working memory capacity (latent variable WMC) to category learning performance for II and RB tasks in Experiment 1. All loadings and correlations are standardized estimates. The correlations between WMC and the two types of task are constrained to be equal. Manifest variables for category learning represent sequence-corrected proportions correct on odd (O) and even (E) trials for rule-based (RB) and information-integration (II) tasks. The difficulty of tasks is coded as easy (EZ), medium (MED), or hard (HD). MU = memory-updating task; OS = operation span task; SSTM = spatial short-term memory task.

of parameters. These BIC values were then converted to BIC weights (Wagenmakers & Farrell, 2004), which approximate the probability that a given participant was using a given strategy for the given problem (see also Lewandowsky & Farrell, 2011).

The evolution of the BIC weights across learning is shown in Figure 5. The figure shows that for the RB tasks, people became increasingly likely across training to use a one-dimensional rule on the appropriate dimension. For the II tasks, information integration increased across training for the two easier tasks (EZ and MED) but remained near the floor for the hardest II task.

Summary statistics for the BIC weights achieved at the end of training for the six different strategies are shown in Table 3. It is clear that for the RB tasks, people predominantly engaged a one-dimensional rule on the appropriate dimension. To a lesser extent, people also engaged an information-integration strategy for the RB tasks. For the II tasks, by contrast, rule-based responding was reduced, and information integration was raised, at least for the two easier tasks. In confirmation of the performance data, for the hardest task (II-HD), people were nearly as likely to use a random response strategy as they were to use any of the other classification strategies combined, and random responding was most likely for that task. Notably, the disjunctive strategy was used hardly at all, and although the conjunctive strategy was the third most popular strategy for the two easier II tasks, its maximum probability of engagement hovered around the 16% mark.

For the remaining analysis of individual differences, we therefore focused only on the rule-based and information-integration strategies. Moreover, for this analysis, we were forced to omit the RB-EZ task because inspection of distributions revealed that all BIC weights for that task straddled an extremely narrow band around .86 for rule use and around .11 for information integration. (The BIC weights for those two models when they are fitted to an idealized perfect response profile based on the actual category boundary are .8904 and .1096, respectively, for rule use and information integration. This suggests that most of our participants honed in on the correct one-dimensional rule and relied entirely on that for responding.)

We represented each task by two indicator variables: One variable represented each participant's summed final BIC weights for the two one-dimensional rules combined, and the other represented the BIC weight for information integration. We then entered those indicator variables for the five tasks (all but RB-EZ) into a structural equation model (raw correlations between all BIC weights and WMC measures are shown in Table 4). The key question to be resolved by this analysis was whether strategy use was uniformly related to WMC or whether only rule use might benefit from greater WMC, as expected in the multiple-systems view.

A first measurement model involved two latent variables, one for each type of task (RB and II), and each connected to the corresponding manifest variables. To obtain a good fit, as suggested by modification indices, the correlations between the error terms of the two strategies (rule use and information integration) were freely estimated for all tasks. That model fit well, $\chi^2(29) = 32.69$, CFI = .981, RMSEA = .034, (90% CI [.0, .084]), SRMR = .0625, and its fit was not reduced when the two latent variables were combined into one, $\chi^2(30) = 32.904$, $\Delta\chi^2(1) = 0.214$. We therefore adopted the single-factor measurement model for strategy use, CFI = .985, RMSEA = .030 (90% CI [.0, .081]), SRMR = .0605.[1]

This measurement model was combined with the earlier WMC measurement model to yield a structural model that fit very well, $\chi^2(59) = 61.983$, $p > .1$, CFI = .986, RMSEA = .022 (90% CI [.0, .064]), SRMR = .0659. This model was further augmented by setting to zero the loadings of the manifest variables for the information-integration BIC weights for the two RB tasks because they were nonsignificant ($p > .08$). When this is done, all remaining weights are significant, and the final model shown in Figure 6 fit well: $\chi^2(59) = 65.551$, $p > .1$, CFI = .978, RMSEA = .026 (90% CI [.0, .066]), SRMR = .0741.

[1] Three of the loadings for the RB tasks were only marginally significant, $.05 < p < .08$, and the fourth loading for information integration for the RB-HD task failed to approach significance; $z = .659$, $p > .10$.

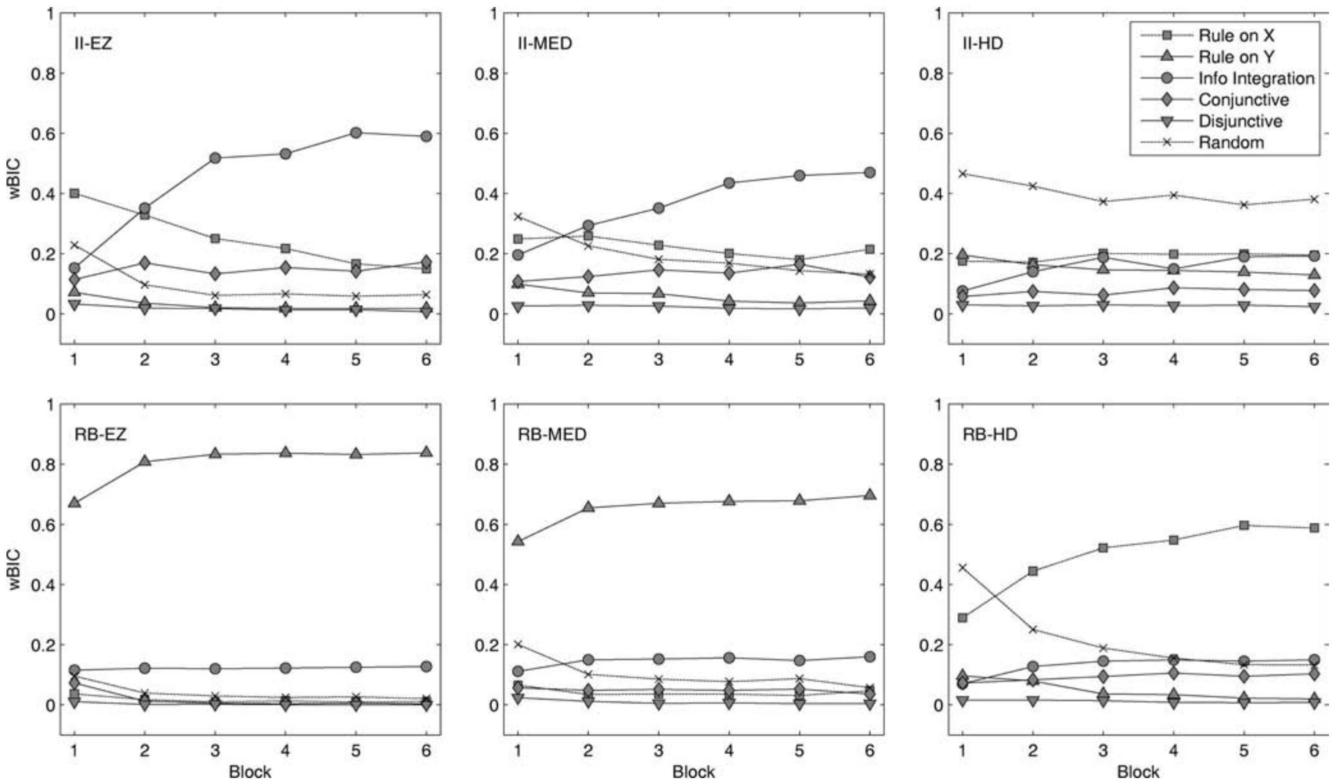

Figure 5. Average Bayesian information criterion weights (wBIC) for the six response models across training blocks in Experiment 1 for all six conditions. Each panel represents one category-learning condition labeled as rule based (RB) or information integration (II). Difficulty was easy (EZ), medium (MED), or hard (HD). The final BIC weights at the end of training are shown in Table 3 and form the basis for the individual-differences analysis. See text for details.

To facilitate interpretation of the figure, it helps to point out that the manifest variables identify the task by the prefixes RB and II, respectively, whereas strategies are identified by the suffix 1D for rule use and II for information integration, respectively. The figure permits several interesting conclusions: First, there was a systematic tradeoff between rule use and information integration, which was particularly pronounced for the II tasks: The more people used information integration, the less likely they were to use rules. This was captured by the negative correlations between error terms for the pairs of alternative strategies and, for the II tasks, also by the negative loadings of the rule use manifest variables (i.e., those ending in 1D) accompanied by positive loadings for information integration. This tradeoff implies that the more likely people were to use information integration, the less likely they were to use a rule to solve the II tasks.

Second, there was a strong correlation between the two latent variables for WMC and strategy use, $r = .581$, $z = 3.516$, $p < .001$, suggesting that WMC was strongly and equally associated with rule use and information integration; whichever strategy was most task-appropriate, people used to an extent predicted by their WMC.

Finally, when rule use was not appropriate for the task at hand, namely in the II tasks, WMC predicted that people would use it less. Thus, rule use was both positively and negatively associated

Table 3
Final-Block Mean (and Standard Error) Bayesian Information Criterion Weights for All Six Strategies Observed in Experiment 1

Task	Rule on *X*	Rule on *Y*	II strategy	Conjunctive	Disjunctive	Random
RB-EZ	.003 (.003)	.861 (.008)	.129 (.004)	.002 (.001)	.000 (.000)	.004 (.004)
RB-MED	.022 (.010)	.749 (.021)	.159 (.010)	.034 (.006)	.004 (.002)	.031 (.012)
RB-HD	.627 (.022)	.015 (.004)	.159 (.013)	.106 (.009)	.007 (.002)	.085 (.020)
II-EZ	.134 (.022)	.012 (.004)	.606 (.036)	.197 (.028)	.006 (.001)	.046 (.012)
II-MED	.204 (.023)	.034 (.005)	.486 (.035)	.132 (.020)	.019 (.003)	.124 (.018)
II-HD	.217 (.019)	.101 (.011)	.203 (.025)	.073 (.010)	.027 (.004)	.379 (.028)

Note. *X* = frequency of Gabor patch; *Y* = orientation of Gabor patch; II = information integration; RB = rule based; EZ = easy; MED = medium; HD = hard.

Table 4
Correlations Between Working Memory Capacity Tasks and Bayesian Information Criterion Weights for Rule Use and Information Integration in Experiment 1

Task	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MU	—												
2. OS	.199*	—											
3. SSTM	.233*	.222*	—										
4. RB-MED-1D	.125	.045	.066	—									
5. RB-MED-II	−.001	.036	.033	−.454**	—								
6. RB-HD-1D	−.066	.232*	−.048	.126	.077	—							
7. RB-HD-II	.152	−.106	.031	−.030	−.087	−.242*	—						
8. II-EZ-1D	.019	−.149	−.065	−.073	−.143	−.201*	.140	—					
9. II-EZ-II	.174	.139	.181	.158	.191*	.108	.000	−.591**	—				
10. II-MED-1D	−.165	−.107	−.114	−.145	−.161	−.119	−.111	.255**	−.247**	—			
11. II-MED-II	.188	.195*	.218*	.018	.164	.120	.175	−.198*	.285**	−.698**	—		
12. II-HD-1D	−.032	.035	−.258**	−.089	.107	−.005	−.122	.074	−.066	.266**	−.166	—	
13. II-HD-II	.039	.036	.184	.136	−.051	.017	.041	−.150	.140	−.269**	.273**	−.345**	—

Note. MU = memory-updating task; OS = operation span task; SSTM = spatial short-term memory task; the prefix RB = rule-based task; the prefix II = information-integration task; the suffix 1D = rule-use strategy; the suffix II = information-integration strategy; MED = medium; HD = hard; EZ = easy.
* $p < .05$. ** $p < .01$.

with WMC, depending on whether a one-dimensional rule was optimal for the task at hand.

The positive association between information integration and WMC runs counter to the expectations of the MMS view. To provide a direct test of the view's expectation that WMC should not be involved in information integration, we ran another model in which the loadings for the information-integration manifest variables for the II tasks (i.e., II-EZ-II; II-MED-II; and II-HD-II) were set to zero. This model fit significantly worse, $\Delta\chi^2(5) = 31.174$, $p < .0001$.

Discussion

The empirical conclusions of Experiment 1 are quite straightforward: A latent variable representing WMC was uniformly associated with category-learning performance, for both II and RB tasks. Moreover, an SEM model that instantiated the predictions of the MMS view by fixing the correlation between WMC and II performance to zero fit significantly worse than an unconstrained model.

Similarly, a single latent variable captured the extent to which people were likely to use one of the two principal strategies, namely, rule use and information integration. This single latent variable was again associated with WMC, suggesting that working memory uniformly contributes not only to directly measured performance, but also to the extent to which people adopt the task-appropriate strategy. This finding runs counter to the explicit expectations of the MMS view (e.g., Ashby & Maddox, 2005; Filoteo et al., 2010; Maddox et al., 2008).

Before we explore the conceptual issues raised by those results, we present a further study that extended Experiment 1 in two ways: First, the easiest (RB-EZ) and hardest conditions (II-HD) were eliminated in order to focus on the midrange of performance, where the II and RB tasks overlapped during training in the first study (Figure 3). Second, we included transfer blocks throughout training on which response feedback was withheld, to permit examination of the emergence of people's ability to generalize their classification strategies to novel items.

Experiment 2

Method

Participants and categorization stimuli. One hundred and nineteen undergraduate students at National Chengchi University participated in this experiment. The participants were paid NT$100 (≅ US$3.50) for each session (about 1 hr). Every participant completed the two sessions on different days within a 2-week period. Additionally, to encourage use of the optimal categorization boundary, a bonus of NT$50 (≅ US$1.75) was paid if accuracy in any one of the learning blocks exceeded a criterion (explained later). The bonus was not paid more than once for each category-learning task. Across the four tasks (2 types × 2 difficulties), the maximum achievable bonus was thus NT$400 (≅ US$14).

The stimuli in Experiment 2 were again Gabor patches that varied along two dimensions, namely, orientation and spatial frequency. In this experiment, the four conditions (RB-MED, RB-HD, II-EZ, and II-MED) were extended by including an additional 44 novel transfer stimuli. The category structures are shown in Figure 7.

Procedure. WMC was again measured using the battery presented by Lewandowsky et al. (2010). Unlike in Experiment 1, all four WMC tasks were used here: An OS task, an SS task, an MU task, and an SSTM task. The SS task was not used in Experiment 1 and is very similar to the OS task, except that the distracter stimuli were sentences, rather than mathematical equations, that had to be judged as syntactically correct. Because all participants in this experiment were Chinese native speakers, the Chinese mode of the WMC battery was chosen for the SS task (for details, see Lewandowsky et al., 2010).

Each categorization task comprised two types of blocks: Training blocks of 66 trials each and transfer blocks of 44 trials involving stimuli not seen during any of the training blocks. The same 44 transfer items appeared in each transfer block.

Altogether there were eight blocks for each task, with Blocks 2, 5, and 8 being transfer blocks and the remainder being training blocks. Training trials were administered in the same way as in

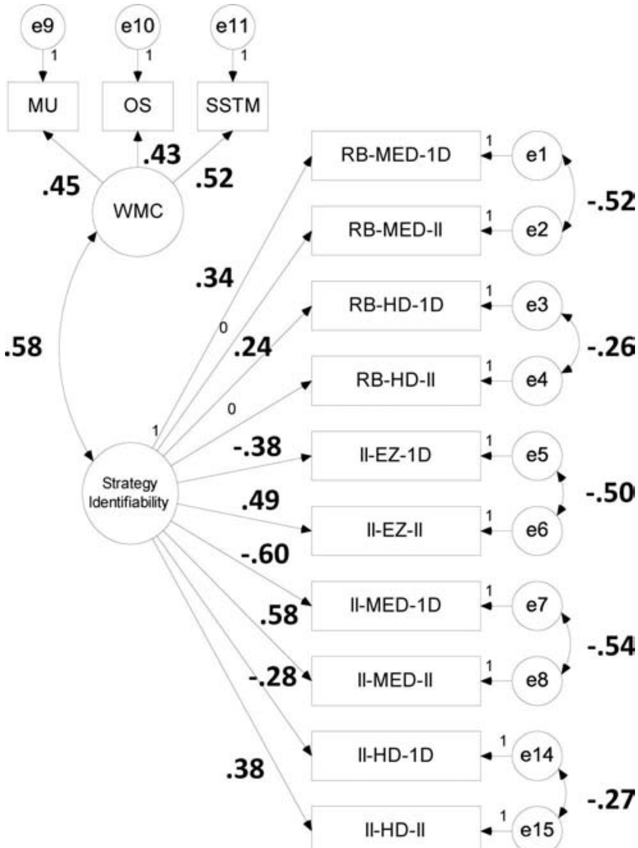

Figure 6. Structural model relating working memory capacity (WMC) to use of categorization strategies in Experiment 1. All loadings and correlations are standardized estimates. Manifest variables for strategy use represent Bayesian information criterion (BIC) weights for rule use (the suffix 1D) or information integration (the suffix II). Prefixes identify rule-based (RB) and information-integration (II) tasks. The difficulty of tasks is coded as easy (EZ), medium (MED), or hard (HD). MU = memory-updating task; OS = operation span task; SSTM = spatial short-term memory task.

Experiment 1. Transfer trials were identical except that no feedback was provided after a response.

Tasks were assigned to the two experimental sessions in two possible sequences. In Sequence 1, the tasks were paired across sessions as follows: RB-HD → OS→ SSTM → II-MED for Session 1 and II-EZ → SS → MU → RB-MED for Session 2. In Sequence 2, the assignment of tasks to sessions was II-EZ → SS → MU → RB-MED for Session 1 and RB-HD → OS → SST → II-MED for Session 2. Participants were randomly assigned to a sequence, with an equal number in each.

Results

Working memory capacity. Summary statistics for the WMC tasks for all participants ($N = 119$) are shown in Table 5, and the corresponding correlations are shown in Table 6. Table 5 also provides estimates of reliability (Cronbach's α), again obtained by comparing thirds of trials for each task.

Categorization. Average performance across blocks is shown in the left panel of Figure 8 for all four tasks. Visual inspection shows that the learning performance in the RB-MED, RB-HD, and II-EZ tasks overlapped considerably, as intended and as expected. The II-MED task led to worse performance ($M = .69$) than the others ($M = .77$). The final learning performance of the II tasks was significantly better than what would be expected from the optimal bilinear alternative; for II-EZ (.85 vs. .76), $t(118) = 8.57$, $p < .0001$, and for II-MED (.76 vs. .73), $t(118) = 2.56$, $p < .05$.

A 2 (task type: II vs. RB) × 2 (difficulty) × 5 (block) within-subjects ANOVA revealed significant main effects for task, $F(1, 118) = 15.30$, $MSE = 0.03$, $p < .01$; difficulty, $F(1, 118) = 36.39$, $MSE = 0.03$, $p < .01$; and block, $F(4, 472) = 294.55$, $MSE = 0.01$, $p < .01$. The two-way interactions involving Task × Difficulty, $F(1, 118) = 31.12$, $MSE = 0.04$, $p < .01$, and Task × Block, $F(4, 472) = 6.18$, $MSE = 0.005$, $p < .01$, were also significant. The two-way interaction involving Difficulty × Block was not significant, $F(4, 472) = 1.88$ $MSE = 0.005$, $p = .11$. Finally, the overarching three-way interaction was not significant, $F(4, 472) = 1.05$, $MSE = 0.005$, $p > .1$. Because focus here is on individual differences and the involvement of working memory, we do not consider those interactions further.

Performance on the transfer blocks was summarized by computing the proportion of correct responses across the three blocks, where each transfer response was considered correct when an item was classified as intended by the experimenter. The resultant means are shown in the right panel of Figure 8 together with the best performance that could be expected on the basis of application of a bilinear boundary.

It is obvious from the figure that people improved across repeated transfer blocks, which was confirmed in a 3 (test block) ×

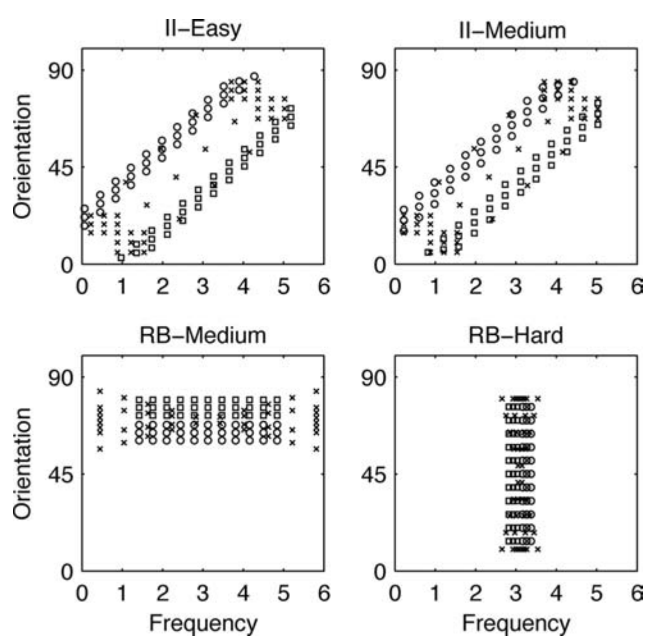

Figure 7. The category structures for all four tasks used in Experiment 2, with information-integration (II) tasks in the top row of panels and rule-based (RB) tasks in the bottom row. Open circles and squares refer, respectively, to the training stimuli of the two categories. The crosses represent transfer stimuli shown during the transfer blocks only and without response feedback.

Table 5
Summary of Working Memory Capacity Scores in Experiment 2

Measure	MU	OS	SS	SSTM
M	0.82	0.79	0.83	0.89
SD	0.14	0.12	0.09	0.14
Minimum	0.30	0.16	0.44	0.77
Maximum	1.00	0.96	1.00	1.00
Skewness	−1.15	−1.63	−1.00	0.05
Kurtosis	1.22	5.05	0.93	−0.13
Cronbach's α	0.80	0.80	0.80	0.93
Standardized loadings	0.64	0.52	0.57	0.47

Note. Standardized loadings refer to the working memory capacity measurement model. MU = memory-updating task; OS = operation span task; SS = sentence span task; SSTM = spatial short-term memory task.

4 (task) within-subjects ANOVA, $F(2, 236) = 148.95$, $MSE = 15.64$, $p < .0001$. Likewise, the obvious differences between tasks were significant, $F(3, 354) = 12.95$, $MSE = 37.78$, $p < .0001$, whereas the interaction between the two variables was not, $F(6, 708) < 1$. The transfer data confirm the pattern evident in the training data, albeit with a much larger corpus of stimuli and in the absence of feedback. Because the transfer data largely confirm the training data, we do not consider the transfer block further and instead focus on exploration of individual differences in the learning data.

Structural equation modeling. A measurement model was again constructed for WMC relying on the four indicator variables (OS, SS, MU, and SSTM). The loadings of the four tasks on the WMC latent variable are shown in Table 5. Following precedent (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Lewandowsky et al., 2010), the correlation between the error terms for the span tasks, SS and OS, was allowed to be freely estimated (.23). Not surprisingly, the model fit extremely well, $\chi^2(1) = 0.317$, $p > .1$, CFI = 1.0, RMSEA = .0 (90% CI [.0, .201]), SRMR = .0103.

For the categorization measurement model, we again used the residuals from a dummy regression that extracted the effect of the counterbalancing sequence. Each task was again represented by two manifest variables with freely estimated correlations between the error terms that represented odd and even trials, respectively. As in Experiment 1, we focused on a measurement model with two latent variables in anticipation of a test of the MMS hypothesis. The model with separate factors for RB and II fit extremely well, $\chi^2(16) = 12.657$, $p > .1$, CFI = 1.0, RMSEA = .0 (90% CI [.0, .067]), SRMR = .0131. As in Experiment 1, the correlations between the residuals within each task were freely estimated, with the exception of the manifest variables for the easier RB task, for which the correlation was not significant. The correlation between the two latent variables was high, $r = .79$, $p < .0001$.

The two measurement models were combined into a structural model involving the three latent variables WMC, RB, and II. To test the predictions of the MMS view, we compared the full unconstrained model, in which all correlations among the three latent variables were freely estimated, $\chi^2(47) = 44.542$, $p > .1$, CFI = 1.0, RMSEA = .0 (90% CI [.0, .055]), SRMR = .0481, to two constrained models. The first constrained model represented the predictions of the MMS view, by setting the correlation between WMC and II to zero. This model fit significantly worse, $\Delta\chi^2(1) = 8.94$, $p < .003$, and although it fit acceptably by some criteria, $\chi^2(48) = 53.484$, $p > .1$, CFI = .995, RMSEA = .031 (90% CI [.0, .070]), its average standardized deviation between imputed and observed values was quite large, SRMR = .105. The second constrained model forced the correlations between WMC and the two category-learning latent variables to be equal. This model fit negligibly worse than the unconstrained model, $\Delta\chi^2(1) = 0.30$, $p > .10$, and fit very well overall, $\chi^2(48) = 44.844$, $p > .1$, CFI = 1.0, RMSEA = .0 (90% CI [.0, .054]), SRMR = .0491. We therefore adopted this model as our final structural model, and it is shown in Figure 9. The figure shows that the two category-learning latent variables were highly correlated, $r = .80$, and that both were related to WMC, $r = .44$, $Z = 3.414$, $p < .0001$.

Response modeling. BIC weights were again computed for each strategy and each participant in the same manner as in Experiment 1. Figure 10 shows the evolution of the BIC weights for the various strategies across training. As before, the figure

Table 6
Correlations Between WMC Tasks and Residualized Category Learning Measures Used in the Structural Equation Modeling for Experiment 2

Task	1	2	3	4	5	6	7	8	9	10	11	12
1. MU	—											
2. OS	.317**	—										
3. SS	.373**	.456**	—									
4. SSTM	.302**	.267**	.252**	—								
5. rRB-MEDe	.236**	.276**	.116	.252**	—							
6. rRB-MEDo	.287**	.288**	.148	.297**	.953**	—						
7. rRB-HDe	.207**	.236**	.129	.176	.456**	.440**	—					
8. rRB-HDo	.205*	.240**	.088	.178	.450**	.422**	.924**	—				
9. rII-EZe	.147	.266**	.081	.234*	.653**	.641**	.282**	.292**	—			
10. rII-EZo	.080	.198*	−.010	.215*	.629**	.640**	.285**	.301**	.888**	—		
11. rII-MEDe	.241**	.230*	.148	.246**	.612**	.612**	.272**	.288**	.639**	.637**	—	
12. rII-MEDo	.185*	.158	.107	.200*	.526**	.517**	.248**	.279**	.622**	.615**	.884**	—

Note. MU = memory-updating task; OS = operation span task; SS = sentence span task; SSTM = spatial short-term memory task; the prefix r = residualized; RB = rule-based task; II = information-integration task; MED = medium; HD = hard; EZ = easy; the suffix e = even-numbered trials; the suffix o = odd-numbered trials.
* $p < .05$. ** $p < .01$.

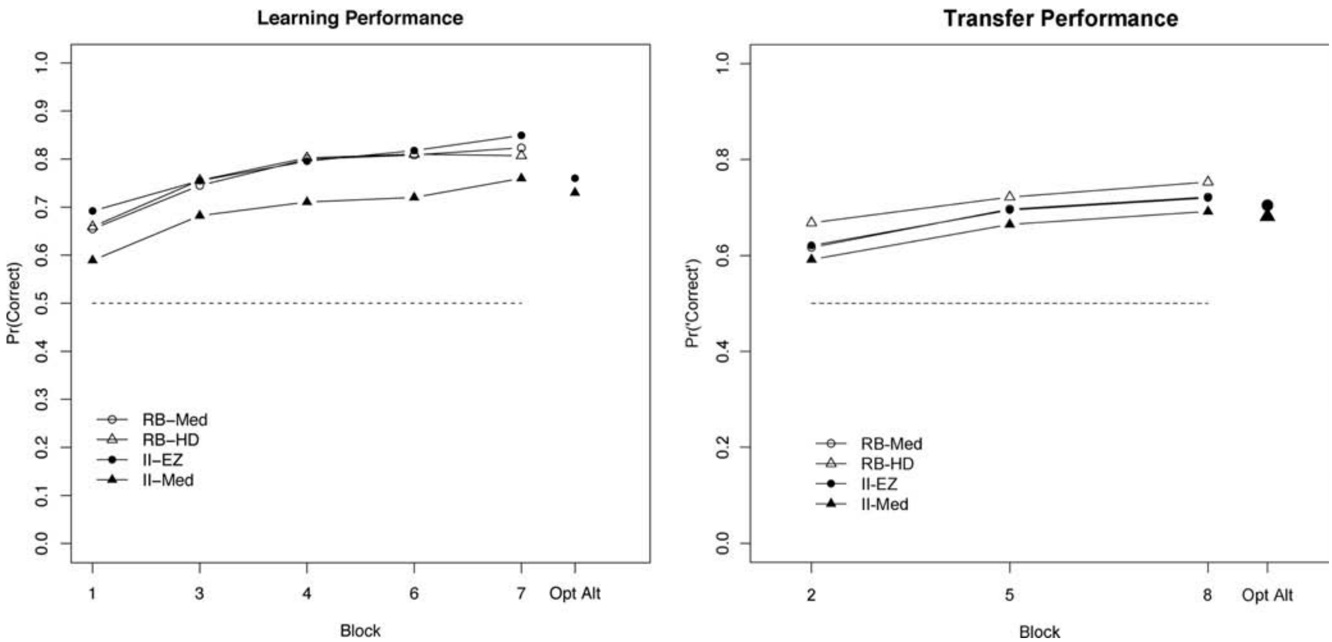

Figure 8. Left panel: Proportion correct across training blocks for all categorization tasks in Experiment 2. Blocks are numbered by their ordinal position in the sequence, with transfer blocks (in Positions 2, 5, and 8) omitted. Filled plotting symbols are for information-integration (II) tasks, and open plotting symbols are for rule-based (RB) tasks. Right panel: Proportion correct across transfer blocks for all categorization tasks in Experiment 2. Blocks are numbered by their ordinal position in the sequence, with training blocks omitted. Filled plotting symbols are for II tasks, and open plotting symbols are for RB tasks. The data points for Opt Alt refer to the best performance that could be expected on the basis of application of two-dimensional bilinear boundaries. Med = medium; HD = hard; EZ = easy.

shows that people gradually shifted toward one-dimensional rule use for the RB tasks and that information integration emerged for the II tasks across blocks, albeit never to the same extent as rule use did for the RB problems.

Summary statistics for the BIC weights achieved at the end of training are shown in Table 7. The table again suggests that for the RB tasks, people predominantly engaged the one-dimensional rule on the appropriate dimension. For the II tasks, by contrast, rule-based responding was reduced and information integration was the modal strategy. For the II-MED task, placing a single-dimensional rule on frequency of the Gabor patch (X) was the favored strategy, but information integration was the second most likely way of classifying those stimuli. The disjunctive strategy was again used hardly at all, and although the conjunctive strategy was the third most popular strategy for the II tasks, its maximum probability of engagement was below 14%. For the individual-differences analysis, we therefore again focused on the rule-based and information-integration strategies only.

The measurement model was constructed in the same manner as for Experiment 1 (raw correlations are shown in Table 8), although in this instance the pairwise correlation between error terms for the two strategies for the RB-EZ task was not freely estimated. A two-factor model, with one latent variable for rule use and another for information integration, fit extremely well, $\chi^2(16) = 10.819$, $p > .1$, CFI = 1.0, RMSEA = .0 (90% CI [.0, .053]), SRMR = .0346, but its fit was not reduced when the two latent variables were combined into one, $\chi^2(17) = 10.899$, $p > .1$, $\Delta\chi^2(1) = 0.08$. We therefore adopted the single-factor measurement model, CFI = 1.0, RMSEA = .0 (90% CI [.0, .035]), SRMR = .0352.[2]

The structural model fit extremely well, $\chi^2(49) = 38.175$, CFI = 1.0, RMSEA = .0 (90% CI [.0, .032]), SRMR = .055. Because the loading of the RB-HD-II manifest variable failed to reach significance, $z = 1.1$, $p > .1$, it was set to zero, and the final model is shown in Figure 11. This final model also fit extremely well, $\chi^2(50) = 39.335$, CFI = 1.0, RMSEA = .0 (90% CI [.0, .033]), SRMR = .0573.

The figure largely reinforces the conclusions drawn from Experiment 1. There was a tradeoff between rule use and information integration for the II tasks, such that increased reliance on information integration was associated with reduced rule use. This was captured both by the negative correlations between error terms (for all but the RB-EZ task) and the negative loadings of the rule-use manifest variables for II tasks.

As in Experiment 1, the correlation between the two latent variables for WMC and strategy use was highly significant, $r = .44$, $z = 3.51$, $p < .0001$. It follows that WMC was strongly associated with both rule use and information integration; people used whichever strategy was most task appropriate to an extent predicted by their WMC. When rule use was not appropriate,

[2] The loading of the RB-HD-II manifest variable failed to reach significance; $z = .936$, $p > .1$.

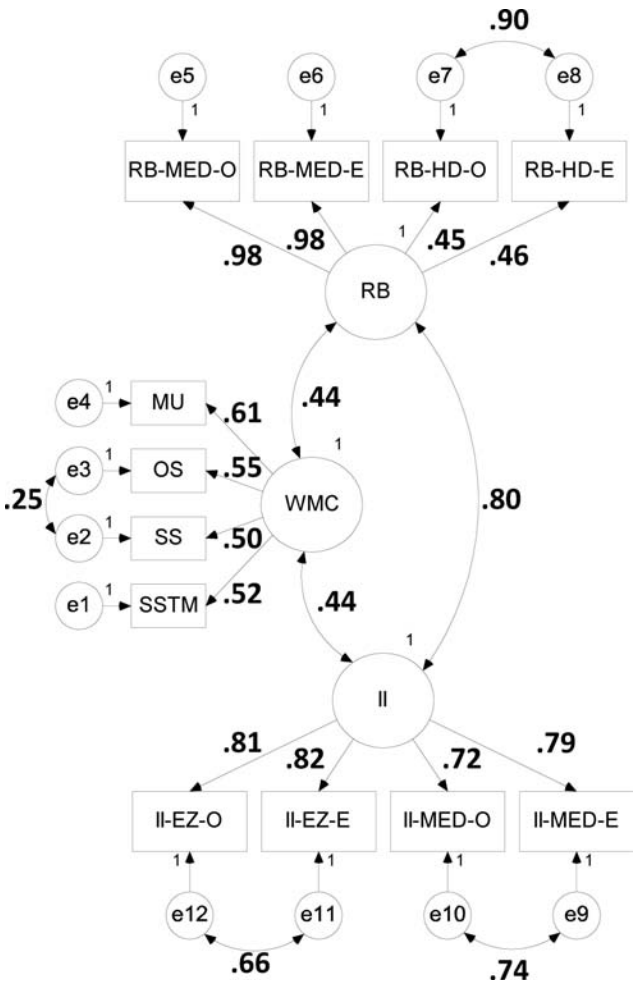

Figure 9. Structural model relating working memory capacity (latent variable WMC) to category learning performance (variables RB and II) in Experiment 2. All loadings and correlations are standardized estimates. Manifest variables for category learning represent sequence-corrected proportions correct on odd (O) and even (E) trials for rule-based (RB) and information-integration (II) tasks. The difficulty of tasks is coded as easy (EZ), medium (MED), or hard (HD). All loadings and coefficients shown are significant, except for the correlation between the residuals for the two span tasks, $p = .062$. MU = memory-updating task; OS = operation span task; SS = sentence span task; SSTM = spatial short-term memory task.

namely, in the II tasks, WMC predicted that people would use it less.

We again tested the specific hypothesis of the MMS view that information integration should not be related to WMC by setting the loadings of all information-integration manifest variables to zero. This model fit dramatically worse, $\Delta\chi^2(3) = 54.84$, thus clearly disconfirming the expectation of the MMS view.

Discussion

Experiment 2 replicated the first study almost exactly: Individual variation in performance on all four tasks was equally associated with WMC. A model in which the correlation between WMC and II task performance was fixed at zero, as predicted by the MMS view, fit significantly worse than the unconstrained model (which, in turn, fit only negligibly better than a model in which the correlations between WMC and performance were equal for RB and II tasks). Likewise, people's WMC again predicted the extent to which they settled on a strategy with which to classify the items.

It is nonetheless perhaps conceivable that the mediating role of WMC on category-learning performance was limited to those individuals who relied on rules to perform the information-integration tasks (although this would fail to explain why greater WMC was associated with greater reliance on information integration; see, e.g., Figure 11). To put to rest this possibility, we reanalyzed the results from both experiments together by considering only those participants who were identified as having maximally relied on an information-integration strategy. That is, only those participants were considered whose BIC weight for information integration was greater than any of the five others. Across the two experiments, for these participants the correlation between (raw) II-EZ performance and WMC was .19, $p < .05$ ($N = 147$), and between II-MED and WMC performance, the correlation was .33, $p < .01$ ($N = 136$), confirming that WMC mediated performance even for the subset of individuals whose reliance on information integration was demonstrably greatest. (Note that this correlation relied on averaging the standardized scores on the subset of working memory tasks that was common to both experiments; the lower absolute magnitude of those correlations compared with that observed between the latent variables is therefore not surprising.) This result should help allay concerns that the strong and uniform mediating role of WMC was somehow confined only to people who relied on rules when learning an II task.

General Discussion

We focused on the relationship between working memory and category learning for two main reasons: first, to assist in adjudicating between competing theories of working memory and, second, to shed light on the debate over single and multiple system views of category learning. We take up those two principal issues after we discuss potential limitations and criticisms of our work.

Potential Limitations

Natural variation versus experimentation. The present study exploited naturally occurring variation among individuals to examine the associations between multiple sources of such individual variation. One obvious drawback of this approach is that the data are necessarily correlational, and one might therefore wonder if experimental manipulations would not represent a preferable avenue for research.

We noted at the outset that several studies have examined the involvement of working memory on category learning through a dual-task methodology (Foerde et al., 2007; Miles & Minda, 2011; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). In those studies, people must perform a secondary task during category learning that ostensibly occupies working memory. Although the results from those dual-task methodologies often support the MMS view, with RB task performance being more affected by the presence of a secondary task than II task performance, we already noted that those studies come with their own sets of problems (e.g., tradeoffs between the primary and secondary tasks, secondary

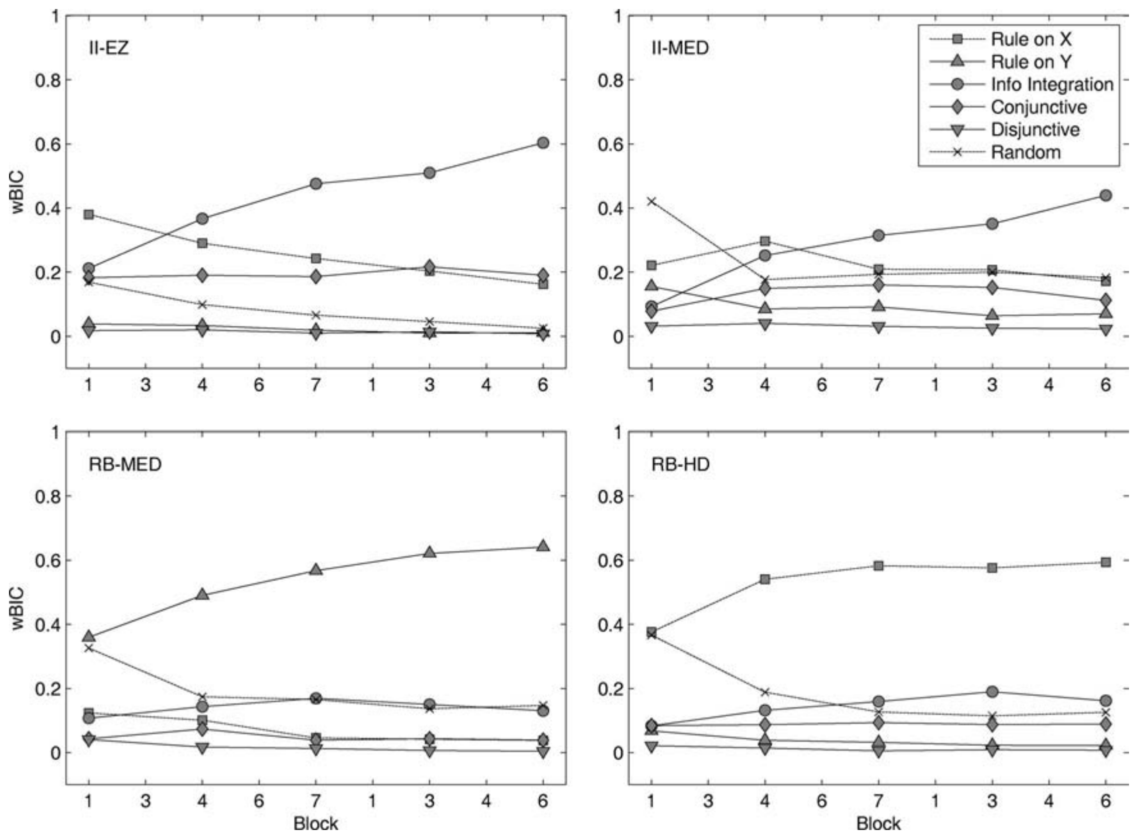

Figure 10. Average Bayesian information criterion weights (wBIC) for the six response models across training blocks in Experiment 2 for all four conditions. Each panel represents one category-learning condition labeled as rule based (RB) or information integration (II). Difficulty was easy (EZ), medium (MED), or hard (HD). The final BIC weights at the end of training are shown in Table 7 and form the basis for the individual-differences analysis. See text for details.

tasks are confined to a single domain). Those problems were resolved here by avoiding the dual-task methodology and by relying on a WMC task battery that was explicitly designed to be domain general (Lewandowsky et al., 2010).

Finally, existing findings from the dual-task paradigm have recently undergone re-examination, and in many instances, the original interpretation has been found wanting. First, Nosofsky and Kruschke (2002) and Nosofsky et al. (2005) showed that the results of Waldron and Ashby's (2001) original study were actually consistent with a single-system model. Second, Zeithamova and Maddox's (2006) study has been shown to be based on a statistical artifact (Newell et al., 2010). Newell et al. (2010) ran three experiments in an attempt to replicate the differential influence of working memory load on the RB task. In their experiments, which differed only in the difficulty of the secondary task, they showed clearly that there was no evidence for any differential influence; the secondary task made both RB and II task learning more difficult in precisely the same way. Newell et al.'s study used a between-subjects design, in which participants were assigned to either an RB or II task consisting of discriminating Gabor-patch figures drawn from two categories. Their stimulus locations were similar to those used in the present study (i.e., stimuli in the RB

Table 7
Final-Block Mean (and Standard Error) Bayesian Information Criterion Weights for All Six Strategies Observed in Experiment 2

Task	Rule on *X*	Rule on *Y*	II strategy	Conjunctive	Disjunctive	Random
RB-MED	.021 (.005)	.667 (.027)	.132 (.009)	.037 (.006)	.004 (.001)	.139 (.027)
RB-HD	.625 (.021)	.018 (.006)	.172 (.014)	.094 (.005)	.008 (.003)	.083 (.020)
II-EZ	.169 (.022)	.011 (.003)	.600 (.032)	.185 (.023)	.008 (.001)	.027 (.008)
II-MED	.169 (.020)	.046 (.008)	.458 (.035)	.113 (.019)	.025 (.007)	.190 (.023)

Note. X = frequency of Gabor patch; Y = orientation of Gabor patch; II = information-integration task; RB = rule-based task; EZ = easy; MED = medium; HD = hard.

Table 8
Correlations Between Working Memory Capacity Tasks and Bayesian Information Criterion Weights for Rule Use and Information Integration in Experiment 2

Task	1	2	3	4	5	6	7	8	9	10	11	12
1. MU	—											
2. OS	.317**	—										
3. SS	.373**	.456**	—									
4. SSTM	.302**	.267**	.252**	—								
5. RB-MED-1D	.160	.220*	.114	.278**	—							
6. RB-MED-II	.105	.139	.009	.106	.282**	—						
7. RB-HD-1D	.135	.020	−.031	.040	.211*	.205*	—					
8. RB-HD-II	.011	.115	.018	.142	.146	.076	−.334**	—				
9. II-EZ-1D	−.102	−.240*	−.006	−.199*	−.402**	−.325**	−.243**	−.011	—			
10. II-EZ-II	.097	.132	−.106	.116	.329**	.271**	.197*	−.054	−.646**	—		
11. II-MED-1D	−.155	−.180	−.077	−.127	−.147	−.151	−.107	.065	.275**	−.172	—	
12. II-MED-II	.113	.152	.143	.166	.307**	.248**	.187*	.040	−.308**	.309**	−.579**	—

Note. MU = memory-updating task; OS = operation span task; SSTM = spatial short-term memory task; SS = sentence span task; the prefix RB = rule-based task; the prefix II = information-integration task; the suffix 1D = rule-use strategy; the suffix II = information-integration strategy; MED = medium; HD = hard; EZ = easy.
* $p < .05$. ** $p < .01$.

task were discriminable along a single dimension, whereas those in the II task required use of both dimensions) and were nearly identical to the items used by Zeithamova and Maddox. Participants were further divided into secondary-task and control (no-secondary-task) groups; the secondary task was manipulated between experiments. In their Experiment 1, participants completed the secondary task used by Zeithamova and Maddox; two digits differing in spatial and numeric magnitude flanked a Gabor patch for 200 ms. The Gabor-patch display was terminated by the participant making their categorization judgment and was followed by 1,000 ms of feedback and 1,000 ms of delay before a prompt appeared cuing participants to respond on the basis of the value or size of the digits. In Newell et al.'s Experiment 2, only the value was prompted. Their Experiment 3 used a different secondary task: A string of five digits appeared for 2,000 ms before being replaced by the Gabor patch, and participants were cued with one of the digits 2,000 ms after making their categorization response; their task was to report the digit that had been presented to the right of the cue. All three of these tasks had a measurable negative effect on learning, but none of them produced any sign of a differential effect on RB over II.

Third, Newell et al. (2007) raised serious questions about Foerde et al.'s (2007) result, which used a multicue probability learning task instead of an II task to index the procedural system, showing that although learning was impaired because of their secondary task, there was no change in the nature of the explicit knowledge participants acquired about their own response strategies. Finally, Miles and Minda's (2011) result stands out in that they actually showed that information-integration learning was impaired by a spatial working memory load.

In summary, although our approach is correlational, it resolves certain problems associated with some experimental precedents. Moreover, a recent critical re-examination of those precedents has revealed converging evidence that supports our main finding, namely, that WMC is uniformly associated with category learning.

WMC: A surrogate for ability? Much work has addressed the relationship between WMC and other high-level cognitive tasks, such as reasoning, and it is clear that WMC is a very strong predictor of reasoning and fluid intelligence (e.g., Kyllonen & Christal, 1990). At first blush, critics might therefore argue that WMC is simply a surrogate for ability and that its association with category learning is, therefore, neither surprising nor particularly interesting.

Multiple arguments speak against this possibility. First, notwithstanding their statistical association, WMC and fluid intelligence represent clearly differentiable constructs. Kane et al. (2005) showed in a meta-analysis that WMC and fluid intelligence share around 50% of their variance, which underscores their association but also clarifies that the two constructs are far from identical or synonymous (see also Ackerman, Beier, & Boyle, 2005).

Second, in support of the unique status of working memory, WMC is not uniformly associated with all higher level cognitive tasks; for example, in recognition memory tasks, WMC is associated only with responses that require conscious recollection but not with those that are based on familiarity alone (Oberauer, 2005). Several other attention-demanding tasks have also been found to be independent of WMC, such as visual search tasks that are revealed to be under conscious control by the presence of set size effects (Kane, Poole, Tuholski, & Engle, 2006). WMC can even be negatively associated with performance in situations in which activation of prior domain-relevant knowledge is misleading (e.g., Ricks, Turley-Ames, & Wiley, 2007).

Finally, and most important, the present studies were designed to test a specific core prediction of the MMS view, namely, that WMC would be uncorrelated with II task performance while being positively associated with RB task performance. As we noted at the outset, this prediction has been a core argument of the MMS approach. The primary neuropsychological model of MMS, COVIS (Ashby et al., 1998, 2011) holds this distinction up as a primary postulate. Ashby et al. (2011) stated, "COVIS postulates two systems that compete throughout learning—an explicit, rule-based system that uses logical reasoning and depends on working memory and executive attention, and an implicit system that uses procedural learning" (p. 65).

The centrality of this claim has been repeated often (e.g., Ashby & Maddox, 2005; DeCaro et al., 2008; Maddox & Ashby, 2004; Maddox et al., 2008). The idea of selective working memory

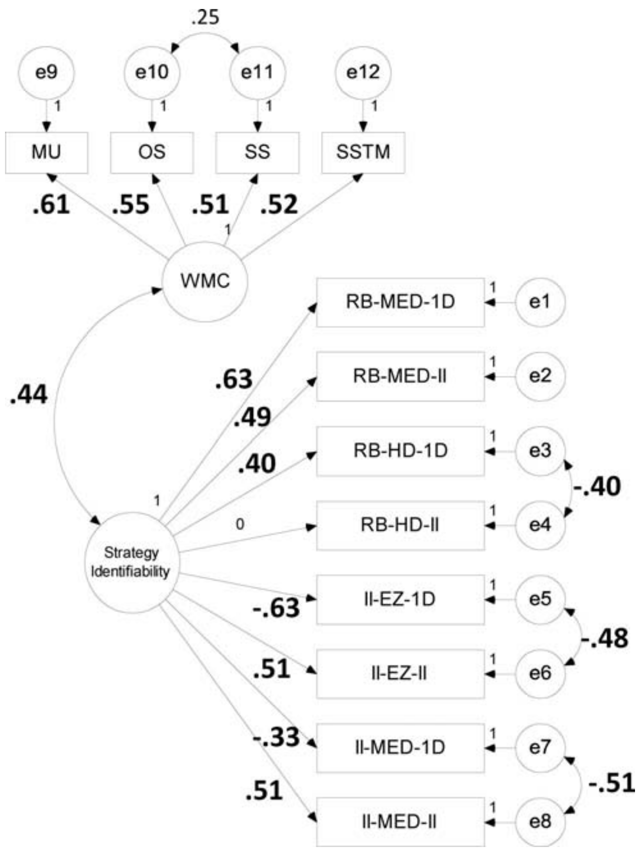

Figure 11. Structural model relating working memory capacity (WMC) to use of categorization strategies in Experiment 2. All loadings and correlations are standardized estimates. Manifest variables for strategy use represent Bayesian information criterion weights for rule use (the suffix 1D) or information integration (the suffix II). Prefixes identify rule-based (RB) and information-integration (II) tasks. The difficulty of tasks is coded as easy (EZ), medium (MED), or hard (HD). MU = memory-updating task; OS = operation span task; SS = sentence span task; SSTM = spatial short-term memory task.

involvement has been used to motivate almost all of the empirical dissociation-based studies taken as evidence for the MMS view (for reviews, see Ashby & Maddox, 2005, 2011; Maddox & Ashby, 2004).

To our knowledge, MMS theoreticians have not appealed to general ability or intelligence in their theorizing at any point. For the purposes of our study, we are thus explicitly unconcerned with fluid intelligence or other potential correlates of working memory, notwithstanding the fact that performance in category learning is likely also mediated by several variables other than WMC. Exploration of such potential candidate variables may present a fruitful avenue for future research; here, we are concerned only with the implications of our data for the MMS view and on theorizing in working memory.

Single Versus Multiple Systems Views of Category Learning

Our data clearly refute a core assumption of multiple-system views of category learning, namely that only the explicit system requires working memory resources, whereas the implicit system does not (e.g., Ashby et al., 1998; Minda & Miles, 2010). In both studies, the relationship between learning performance and WMC was identical for II and RB tasks. A structural equation model that forced WMC to be uncorrelated with II task performance fit significantly worse than a model in which both RB and II task performance were identically associated with WMC.

Importantly, an analysis of individual performance at the level of strategy use buttressed the positive mediating role of WMC. In both studies, a single latent variable captured the extent to which individuals were likely to have used one of two principal strategies: rule use and information integration. This single latent variable was also positively associated with WMC, suggesting a uniform contribution of working memory to the adoption of task-appropriate strategies. This finding is particularly important because it demonstrates that even when we focus on those participants who solved an II task in the manner typically claimed to be under control of a system independent of implicit and working memory, higher WMC nonetheless facilitated strategy adoption and learning. This result meshes well with a study by Craig and Lewandowsky (in press), which also found that WMC predicted how well people learned various category structures but that it did not predict which of several (equally valid) strategies people adopted.

It is worth re-emphasizing the nomenclature provided at the outset: *RB* and *II tasks* refer to physical task parameters, whereas *rule use* and *information integration* refer to the cognitive strategies that people use to solve those categorization tasks. Figures 5 and 10, which plot the evolution of the BIC weights for the principal strategies across blocks, highlight that there is no simple one-to-one match between the physical task parameters and the strategy a participant adopts to solve the task: There is always some probability that a participant will use a rule for an II task and vice versa. However, as Tables 3 and 7 show, by the end of training, higher likelihoods were found for the more task-appropriate strategies. Put simply, this shows that people learn: Reliance on random responding reduces across blocks as reliance on more appropriate strategies increases. Most important, there is no suggestion that this learning should be attributed to separable systems distinguished on the basis of working memory involvement: Whatever the task and however a person approaches it, WMC is uniformly involved in learning and performance.

The generality of that conclusion is underscored by a growing number of converging findings. Evidence for a uniform involvement of WMC has now been obtained across a wide range of categorization tasks. In addition to the Gabor patches used here, Craig and Lewandowsky (in press) found WMC to be associated with learning of the classic 5-4 (Medin & Schaffer, 1978) and correlated-cues (Medin, Altom, Edelson, & Freko, 1982) tasks involving binary dimensions. Sewell and Lewandowsky (in press) found that WMC was associated with learning performance (but not attentional shifts) in blocking and highlighting paradigms (cf. Kruschke, Kappenman, & Hetrick, 2005), and they found a similar positive association between WMC and learning in a complex task involving knowledge partitioning (Yang & Lewandowsky, 2003, 2004) and knowledge restructuring (Sewell & Lewandowsky, 2011). Finally, we already noted Lewandowsky's (2011) finding that WMC supported learning of all six problem types of the classic Shepard tasks. Common to all those experiments is the use of a broad range of tasks to measure WMC, a relatively large sample size, and the use of SEM—all three features are arguably necessary to guard against obtaining results that are difficult to coordinate with precedents or existing theory (e.g., Blair et al., 2009; DeCaro et al., 2008, 2009; Erickson, 2008). The latter class of results is characterized by use of a single task to measure WMC, thus confounding task-specific variance with the construct of interest, and often is also characterized by small sample sizes and use of a single categorization task.

Taken together with recent re-evaluations of data that were initially thought to be supportive of the MMS view (Newell et al., 2010; Nosofsky & Kruschke, 2002; Tharp & Pickering, 2009), the sum total of available evidence now implies fairly strongly that working memory is uniformly and positively associated with category learning performance, regardless of the task structure and irrespective of how people choose to perform the task.

If selective working memory involvement is not a characteristic of the explicit–implicit system distinction, then what other characteristics remain to support the claim for duality? One is the notion that learning is verbalizable in the explicit system but unavailable to awareness in the implicit system. We do not address this particular minefield in detail here; suffice it to say that the evidence for inaccessibility to awareness is often based on questionable measures and that when more sensitive measures are used, relevant explicit knowledge can be revealed even for tasks supposedly under implicit control (e.g., Lagnado, Newell, Kahan, & Shanks, 2006; Newell et al., 2007).

Another major arena of evidence for systems duality is, of course, in neuropsychology and neuroscience. Many authors argue that the explicit and implicit systems are associated with distinct neural pathways and structures (e.g., Ashby et al., 2011) and that this provides convergent evidence for behavioral dissociations observed between implicit and explicit tasks, particularly as instantiated in the II and RB task categorization problems. Although such evidence sometimes appears overwhelmingly supportive of multiple system views (for a review, see Poldrack & Foerde, 2008), there are several reasons to urge caution in the interpretation of the evidence (for a comprehensive critique, see Newell et al., 2011). Summarizing the voluminous literature on this issue is beyond the present scope; however, we can illustrate the need to exercise interpretative caution with two examples. First, although detailed claims about underlying neurobiology have been used to make predictions about the differential effects of variables on performance in II and RB tasks (e.g., delaying feedback; Maddox & Ing, 2005), subsequent neuroimaging work has suggested that the regions crucial to the predictions (i.e., the tail of the caudate nucleus) do not appear to be involved in learning the relevant tasks (i.e., II task; Waldschmidt & Ashby, in press). Moreover, Dunn, Newell, and Kalish (in press) have shown that the dissociation with delay disappears when theoretically irrelevant aspects of the procedure are changed.

Second, recent neuroimaging data from other perceptual category learning tasks (dot-pattern classification) suggest that differences in neural activity observed when participants learn under explicit (intentional) or implicit (incidental) conditions are not necessarily signatures of separable implicit and explicit neural systems. Gureckis, James, and Nosofsky (2011) demonstrated that a dissociation previously interpreted by Reber, Gitelman, Parrish, and Mesulam (2003) as supporting a multiple-system view was more readily interpreted as due to differences in the specific stimulus-encoding instructions given to participants in the implicit and explicit conditions. This re-evaluation emphasizes the need for caution when drawing conclusions about multiple systems on the basis of evidence from neuroimaging.

The need for caution is also highlighted in a recent study by Milton and Pothos (in press), who found considerable overlap in the brain regions involved in a simple and a complex categorization task. The complex category structure they used contained many of the hallmarks of an II task (e.g., a diagonal decision bound), but there was no evidence from the fMRI that structures claimed to underlie II task learning (e.g., the tail of the caudate nucleus) were activated. Rather, both simple and complex tasks recruited areas associated with explicit rule learning (e.g., ventrolateral frontal cortex).

On balance, what is the current status of the multiple-systems view of categorization? In our view, it would be premature, to say the least, to declare the debate closed in favor of multiple systems. On the contrary, concerning the selective involvement of working memory in RB tasks, we find that the sum total of the available data provides no support for this core tenet of the MMS view of category learning. Working memory is clearly involved in all forms of category learning, whether it is based on rules or on information integration. We suggest that this assessment finds support even in recent studies that purport to support the multiple-systems view: For example, Miles and Minda (2011) found that a visuospatial secondary task impaired performance in an II task as well as an RB task, contrary to a dissociation previously reported with a similar task (Zeithamova & Maddox, 2007).[3] This finding is difficult to reconcile with the multiple-systems view, especially in light of the fact that every conceivable effect of a secondary task has now been obtained with each class of tasks: Secondary tasks may or may not affect RB task performance and they may or may not affect II task performance (Miles & Minda, 2011; Newell et al., 2010; Zeithamova & Maddox, 2007). Our data thus contribute to the continued re-evaluation of the prevalence of the multiple-systems perspective in the category learning literature (e.g., Newell et al., 2011).

That said, it would be equally premature to argue that the multiple-systems view is no longer empirically supported. On the contrary, recent evidence from nonhuman species (Smith et al., 2011), reports of dissociations based on mood (Nadler, Rabi, & Minda, 2010), and developmental data (Huang-Pollock, Maddox, & Karalunas, 2011) all point to the resilience of the multiple-systems view. It is therefore possible that a more nuanced version of the multiple-systems view, which acknowledges that working memory is involved in both rule use and information integration, will ultimately prevail.

Role of Working Memory in Category Learning

Turning to the role of working memory in category learning, we first note that there is much evidence in other domains that work-

[3] Miles and Minda (2011) additionally found that a verbal secondary task disrupted RB but not II task performance, which lends support to the distinction between memory systems. However, the robustness of that outcome is called into question by the fact that sometimes a verbal secondary task—unlike a visuospatial task—does not affect RB performance (Zeithamova & Maddox, 2007).

ing memory is closely related to retention and learning over the long term. For example, Mogle, Lovett, Stawski, and Sliwinski (2008) recently argued that the long-term memory involvement in working memory tasks was responsible for the known strong relationship between WMC and fluid intelligence. Similarly, Unsworth and Engle (2007) showed that WMC predicts performance in measures of recollection that are typically taken to reflect long-term memory involvement.

How, then, does WMC support category learning? At least two alternative mechanisms can be cited: On the one hand, WMC could facilitate memory for specific exemplars, such that people with high WMC are able to form more lasting or more exact memories of instances. On the other hand, WMC might facilitate faster learning of some other task-relevant representation (e.g., a category boundary). Several lines of evidence favor the second possibility over the first one, especially for the stimuli being used in our experiments. First, in a thorough comparison of exemplar models and rule-based theories of perceptual categorization, Rouder and Ratcliff (2004) showed that for stimuli and tasks that were related to ours, people are unlikely to rely on exemplar memory. In their study, exemplar models tended to characterize performance only when there were few and highly discriminable stimuli. Given that our stimuli were neither few in number nor highly discriminable, exemplar memory is less likely to have been the principal learning mechanism. Second, even when performance can be characterized by exemplar models, as in the study of the Shepard tasks by Lewandowsky (2011), individual differences turn out not to be related to the precision of exemplar memory. Lewandowsky showed that the specificity parameter within an exemplar theory (ALCOVE; Kruschke, 1992) was unrelated to WMC. That parameter governs the precision of exemplar memories, with greater precision supporting better performance in learning tasks—such as the Shepard problems—that do not require generalization. The fact that the differences in individual sensitivity-parameter estimates did not correlate across tasks or with WMC suggests that exemplar memory was unlikely to be the principal driver of performance. We therefore tentatively endorse the second option, namely, that WMC facilitates speed of learning of whatever representation underlies categorization when exemplar memory is likely to be secondary. In support, Lewandowsky showed that WMC was related to the speed with which weights were updated in two neural-network models (one of which did not involve memory for exemplars). As in the present studies, a single latent variable captured the individual variation of those parameter values across all Shepard tasks: That is, the independent learning-parameter estimates for the six tasks loaded onto a single latent variable, which in turn was associated with WMC. This result is entirely consonant with our present finding that whatever strategy is most task appropriate is adopted in a manner that is uniformly mediated by WMC.

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

A crucial property of the direct-access region is that it temporarily binds together representations that are required for cognitive operations. For example, item representations may be bound to their temporal context, they may be bound to a spatial location, and they may be transformed before being bound to a new or different context (e.g., Ecker et al., 2010). The notion of binding is particularly relevant in the present context because long-term learning is thought to involve transfer of information from the direct-access region to long-term memory (Oberauer, 2009, elaborated on the presumed transfer process and additionally noted the importance of transformation of the information, from a temporary relational format to a unitized or chunked structure.) In support, recent research has uncovered further linkages between performance in complex-span tasks and subsequent long-term retention on surprise final tests (Loaiza & McCabe, in press; Loaiza, McCabe, & Youngblood, in press).

In summary, we offer a rather unequivocal empirical contribution: WMC is strongly related to the two principal manifestations of perceptual category learning, rule use and information integration, both in terms of overall performance increment and strategy adoption. We find no evidence for a dissociation on the basis of working memory involvement between RB and II tasks and their presumed underlying memory systems. Our data form a distinct benchmark for further theorizing that relates categorization and working memory, two acknowledged pillars of human cognition. At a theoretical level, our main conclusion is that we failed to find evidence for a key prediction of the multiple-systems view of categorization, namely, selective involvement of working memory in RB tasks. Importantly, this conclusion rests not only on average performance levels for the II tasks, but also on the fact that WMC was associated with the extent to which people adopted an information-integration strategy.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30–60. doi:10.1037/0033-2909.131.1.30

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442–481. doi:10.1037/0033-295X.105.3.442

Ashby, F. G., Ell, S. E., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition, 31,* 1114–1125. doi:10.3758/BF03196132

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149–178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences, 1224,* 147–161. doi:10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences, 9,* 83–89. doi:10.1016/j.tics.2004.12.003

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 65–87). New York, NY: Cambridge University Press.

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research, 40,* 235–259. doi:10.1207/s15327906mbr4002_4

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. doi:10.3758/BF03196772

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

DeCaro, M. S., Carlson, K. D., Thomas, R. D., & Beilock, S. L. (2009). When and how less is more: Reply to Tharp and Pickering. *Cognition, 111,* 415–421. doi:10.1016/j.cognition.2009.03.001

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition, 107,* 284–294. doi:10.1016/j.cognition.2007.07.001

Dunn, J. C., Newell, B. R., & Kalish, M. L. (in press). The effect of feedback delay and feedback type on perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 170–189. doi:10.1037/a0017891

Erickson, M. A. (2008). Executive attention and task switching in category learning: Evidence for stimulus-dependent representation. *Memory & Cognition, 36,* 749–761. doi:10.3758/MC.36.4.749

Estes, W. K. (1994). *Classification and cognition.* New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195073355.001.0001

Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science, 21,* 415–423. doi:10.1177/0956797610362646

Foerde, K., Poldrack, R. A., & Knowlton, B. J. (2007). Secondary-task effects on classification learning. *Memory & Cognition, 35,* 864–874. doi:10.3758/BF03193461

Gureckis, T. M., James, T. W., & Nosofsky, R. M. (2011). Re-evaluating dissociations between implicit and explicit category learning: An event related fMRI study. *Journal of Cognitive Neuroscience, 23,* 1697–1709. doi:10.1162/jocn.2010.21538

Huang-Pollock, C. L., Maddox, W. T., & Karalunas, S. L. (2011). Development of implicit and explicit category learning. *Journal of Experimental Child Psychology, 109,* 321–335. doi:10.1016/j.jecp.2011.02.002

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. doi:10.1037/0096-3445.130.2.169

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 66–71. doi:10.1037/0033-2909.131.1.66

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of "executive attention". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 749–777. doi:10.1037/0278-7393.32.4.749

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence, 37,* 374–382. doi:10.1016/j.intell.2009.03.004

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

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1,* 106–120.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44. doi:10.1037/0033-295X.99.1.22

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 830–845. doi:10.1037/0278-7393.31.5.830

Kyllonen, P., & Christal, R. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14,* 389–433.

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple cue learning. *Journal of Experimental Psychology: General, 135,* 162–183. doi:10.1037/0096-3445.135.2.162

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 720–738. doi:10.1037/a0022639

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice.* Thousand Oaks, CA: Sage.

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. doi:10.3758/BRM.42.2.571

Lewandowsky, S., Roberts, L., & Yang, L.-X. (2006). Knowledge partitioning in categorization: Boundary conditions. *Memory & Cognition, 34,* 1676–1688. doi:10.3758/BF03195930

Loaiza, V. M., & McCabe, D. P. (in press). Temporal–contextual processing in working memory: Evidence from delayed cued recall and delayed free recall tests. *Memory & Cognition.*

Loaiza, V. M., McCabe, D. P., & Youngblood, J. L. (in press). The influence of levels of processing on recall from working memory and delayed recall tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109,* 35–54. doi:10.1037/0033-295X.109.1.35

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes, 66,* 309–332. doi:10.1016/j.beproc.2004.03.011

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 650–662. doi:10.1037/0278-7393.29.4.650

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 100–107. doi:10.1037/0278-7393.31.1.100

Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition, 108,* 578–589. doi:10.1016/j.cognition.2008.03.010

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science, 17,* 944–948. doi:10.1111/j.1467-9280.2006.01809.x

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7,* 64–82. doi:10.1037/1082-989X.7.1.64

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 37–50. doi:10.1037/0278-7393.8.1.37

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238. doi:10.1037/0033-295X.85.3.207

Miles, S. J., & Minda, J. P. (2011). The effects of concurrent verbal and visual tasks on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 588–607. doi:10.1037/a0022309

Milton, F., & Pothos, E. M. (in press). Category structure and the two learning systems of COVIS. *European Journal of Neuroscience.*

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1518–1533. doi:10.1037/a0013355

Minda, J. P., & Miles, S. J. (2010). The influence of verbal and nonverbal processing on category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 52, pp. 117–162). San Diego, CA: Academic Press.

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? *Psychological Science, 19,* 1071–1077. doi:10.1111/j.1467-9280.2008.02202.x

Nadler, R. T., Rabi, R., & Minda, J. P. (2010). Better mood and better performance: Learning rule-described categories is enhanced by positive mood. *Psychological Science, 21,* 1770–1776. doi:10.1177/0956797610387441

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Science, 12,* 285–290. doi:10.1016/j.tics.2008.04.009

Newell, B. R., Dunn, J. C., & Kalish, M. L. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition, 38,* 563–581. doi:10.3758/MC.38.5.563

Newell, B. R., Dunn, J. C., & Kalish, M. L. (2011). Systems of category learning: Fact or fantasy? In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 167–215). San Diego, CA: Academic Press.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review, 14,* 505–511. doi:10.3758/BF03194098

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review, 9,* 169–174. doi:10.3758/BF03196274

Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition, 33,* 1256–1271. doi:10.3758/BF03193227

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134,* 368–387. doi:10.1037/0096-3445.134.3.368

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 51, pp. 45–100). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(09)51002-X

Oberauer, K., Süβ, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences, 29,* 1017–1045. doi:10.1016/S0191-8869(99)00251-2

Oberauer, K., Süβ, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York, NY: Oxford University Press.

Oberauer, K., Süβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31,* 167–193. doi:10.1016/S0160-2896(02)00115-0

Pelli, D. G. (1997). The video toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. doi:10.1163/156856897X00366

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews, 32,* 197–205. doi:10.1016/j.neubiorev.2007.07.007

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience, 15,* 574–583. doi:10.1162/089892903321662958

Ricks, T. R., Turley-Ames, K. J., & Wiley, J. (2007). Effects of working memory capacity on mental set due to domain knowledge. *Memory & Cognition, 35,* 1456–1462. doi:10.3758/BF03193615

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General, 133,* 63–82. doi:10.1037/0096-3445.133.1.63

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. doi:10.1214/aos/1176344136

Sewell, D. K., & Lewandowsky, S. (2011). Restructuring partitioned knowledge: The role of recoordination in category learning. *Cognitive Psychology, 62,* 81–122. doi:10.1016/j.cogpsych.2010.09.003

Sewell, D. K., & Lewandowsky, S. (in press). Attention and working memory capacity: Insights from blocking, highlighting, and knowledge restructuring. *Journal of Experimental Psychology: General.*

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

Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review, 18,* 414–421. doi:10.3758/s13423-010-0047-8

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,* 411–415. doi:10.1016/j.cognition.2008.10.003

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1,* 31–65. doi:10.1146/annurev.clinpsy.1.102803.144239

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132. doi:10.1037/0033-295X.114.1.104

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192–196. doi:10.3758/BF03206482

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review, 8,* 168–176. doi:10.3758/BF03196154

Waldschmidt, J. G., & Ashby, F. G. (in press). Cortical and striatal contributions to automaticity in information-integration categorization. *NeuroImage.*

Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology, 18,* 18–50. doi:10.1080/09541440500215921

Yang, L.-X., & Lewandowsky, S. (2003). Context-gated knowledge partitioning in categorization. *Journal of Experimental Psychology: Learn-*

ing, Memory, and Cognition, 29, 663–679. doi:10.1037/0278-7393.29.4.663

Yang, L.-X., & Lewandowsky, S. (2004). Knowledge partitioning in categorization: Constraints on exemplar models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1045–1064. doi:10.1037/0278-7393.30.5.1045

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition, 34,* 387–398. doi:10.3758/BF03193416

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition, 35,* 1380–1398. doi:10.3758/BF03193609

Appendix

Definition of Response Models

The aim of our response surface modeling was to characterize each participant's responses at each block of our experiment in a maximally diagnostic manner. The multiple-systems approach traditionally distinguishes between response surfaces that result from the use of readily verbalizable rules and those that reflect information integration (e.g., Maddox & Ashby, 2004; Maddox, Ashby, & Bohil, 2003; Zeithamova & Maddox, 2006). To cover an even broader range of strategy alternatives, we designed six response models. We chose the framework of naïve Bayes classifiers to express our response surfaces; five of those models turned out to be reparameterizations of the models used by Maddox et al. (2003) and others. The sixth is a biased random-guessing model.

The six response models represent two different kinds of rule-based response surfaces, one information-integration strategy and a random-guessing approach. Figure A1 illustrates the classification strategies for all but the random-guessing model. The models

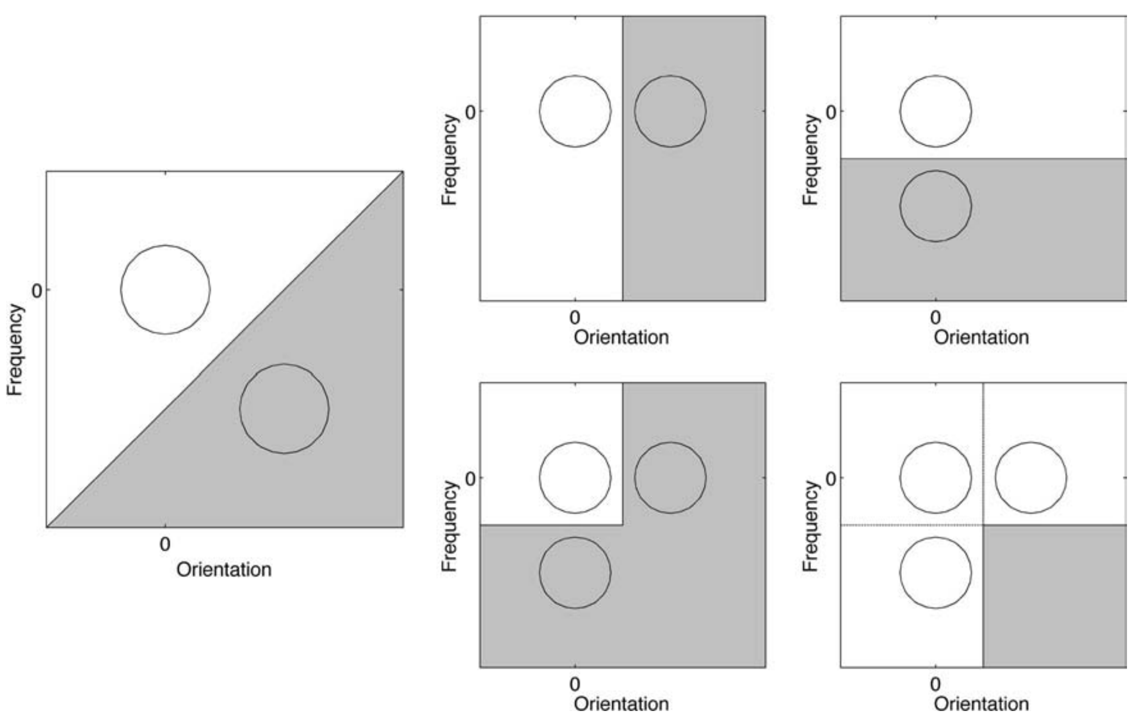

Figure A1. Candidate response models for both experiments. Each panel shows the category space formed by the orientation and frequency dimensions of the Gabor stimuli. Within each panel, the shaded area represents classification responses for one category, and the unshaded area represents responses that fell into the other category. In fact, the transitions from one category to the other form a smooth gradient rather than the binary boundaries shown here for illustrative purposes only. The circles in each panel represent illustrative equiprobability contours of the estimated response distributions. Left panel: information integration; top panels: one-dimensional rule use; bottom panels: conjunctive and disjunctive two-dimensional rule use, respectively.

(*Appendix follows*)

are all based on Bayes's rule for the probability of classifying an item as a member of Category A as given by the following ratio:

$$p(A|x)=\frac{p(x|A)p(A)}{p(x|A)p(A)+p(x|B)p(B)}.$$

The first five models hold that $p(A) = p(B)$, and that the likelihood $p(x|A)$ is given by a bivariate Gaussian distribution with mean $\mu_{\mathbf{A}}$ and standard deviation σ. For convenience, we set $\mu_{\mathbf{A}}$ equal to zero by subtracting the mean of the stimuli given label A by the subject from the value of each stimulus. The five models differed by using different representations of $p(x|B)$, as follows.

The general linear classifier defines likelihood as a bivariate normal with a freely estimated mean $\mu_{\mathbf{B}}$ and a common standard deviation σ and so has three parameters (see left panel in Figure A1). This produces a posterior with an equiprobability boundary that forms a straight line at some location and with some orientation determined by $\mu_{\mathbf{B}}$ and with a gradient determined by the relationship between σ and the magnitude of $\mu_{\mathbf{B}}$.

When the mean of B is set to zero on one of the two dimensions, the resulting posterior has a equiprobability boundary that forms a straight line parallel to the nonzero dimension. The location is determined by μ_B and the gradient by σ. These models are consistent with one-dimensional rules, as shown by the two models in the top row of Figure A1.

It is possible to use the same formalism to construct models consistent with two-dimensional conjunctive or disjunctive rules as well. These are shown in the bottom row of Figure A1. The likelihood for the conjunctive model is formed by setting

$$p(x|B) = \max[p(x|B_1), p(x|B_2)],$$

where $p(x|B_1)$ is given by a multivariate Gaussian with mean (B_1, 0) and standard deviation σ. Similarly $p(x|B_2)$ is given by a distribution with mean (0, B_2). The likelihood for the disjunctive model is just $\min[p(x|B_1), p(x|B_2)]$. The values of B_1 and B_2 determine the locations of the equiprobability boundaries in these models, and the gradient across the boundary is again a function of σ.

Finally, a random model (not shown in Figure A1) assumed A and B were uniform distributions across the entire stimulus space but that $p(A)$ could vary from zero to one to represent a simple biased guessing strategy; $p(B) = 1 - p(A)$.

All parameters were estimated for each block and for each participant–task cell by maximizing the likelihood of the person's observed response profile. For the guessing model, its only parameter was determined simply by observing the relative frequency of A and B responses. For each model, the best fit attained across several runs with different starting values was used to compute BIC weights for the analysis.

Received June 2, 2011
Revision received December 7, 2011
Accepted January 2, 2012 ■