Journal of Experimental Psychology:
Learning, Memory, and Cognition
2017, Vol. 43, No. 4, 503–514


0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000323

# More Is Generally Better: Higher Working Memory Capacity Does Not Impair Perceptual Category Learning

Michael L. Kalish
Syracuse University

Ben R. Newell
University of New South Wales

John C. Dunn
University of Adelaide

It is sometimes supposed that category learning involves competing explicit and procedural systems, with only the former reliant on working memory capacity (WMC). In 2 experiments participants were trained for 3 blocks on both filtering (often said to be learned explicitly) and condensation (often said to be learned procedurally) category structures. Both experiments (total $N = 160$) demonstrated that participants with higher WMC tended to be more accurate in condensation tasks, but not less accurate in filtering tasks. Furthermore, state-trace analysis did not find a differential influence of WMC on performance in these tasks. Finally, inspection of the mixture of response strategies at play across the 2 conditions and 3 blocks showed only a minor influence of WMC, and then only on later training blocks. The results provide no support for the existence of a "system" of category learning that is independent of working memory and are instead consistent with most single-system interpretations of category learning.

*Keywords:* category learning, categorization, working memory, dissociable systems, implicit learning



Categorization is a ubiquitous cognitive challenge. Even considering only those categorical discriminations that depend on perceptual features, there is a vast range of different kinds of categories to be learned, from the names of polygons to the sexes of chicks (Horsey, 2002), The flexibility and the variability in the ways people learn new categorical discriminations are often thought to require the operation of at least two independent processing modes or systems (e.g., Ashby, Paul, & Maddox, 2011; Minda & Miles, 2010). This account, exemplified by the COVIS computational model (Ashby et al., 2011), highlights a distinction between a system that allows people to learn via simple rules, and a second system that allows people to learn through stimulus–response associations. This distinction is claimed to capture properties of real world categories where some things can be classified with verbal rules (e.g., squares vs. triangles) but others, sometimes, cannot (e.g., whether an X-ray depicts a tumor; Ashby & Maddox, 2005). However, many have argued that the flexibility and variety of categorization do not necessitate a commitment to multiple distinct processing systems (e.g., Dunn, Newell, & Kalish, 2012; Lewandowsky, 2011; Newell & Dunn, 2008; Newell, Dunn, & Kalish, 2010, 2011; Newell, Lagnado, & Shanks, 2007; Nosofsky & Zaki, 1998; Palmeri & Flanery, 2002; Speekenbrink, Channon, & Shanks, 2008). Distinguishing one system from two is always difficult, and depends critically on the properties of the proposed multiple systems (Dunn, Kalish, & Newell, 2014).

In the case of perceptual categorization, COVIS proposes that a central feature of the rule-based system is its dependence on working memory capacity (WMC). Rules are said to be held in declarative memory, so that people with higher WMC are better able to acquire, maintain, and switch between categorization rules. For COVIS, this is the fundamental quality of the rule-based system that discriminates it from a proposed procedural system, which operates implicitly, without awareness, and independently of any WMC constraints. The current view of the procedural system is that it learns by associating category labels with exemplar-centered regions of perceptual space (Ashby et al., 2011).

The original impetus for the present experiments is a set of counterintuitive results reported by DeCaro, Thomas, and Beilock (2008). In their experiment, participants learned one of two categorization tasks, defined over a common type of stimulus. Participants could solve one kind of task, which DeCaro et al. called "rule-based," by attending to the values of just one stimulus dimension. The other, which they called "information-integration," required use of multiple dimensions. Because these names prejudge the nature of the classification strategies people might use to solve the respective tasks, we use an older terminology for the single versus multiple dimension distinction: filtering versus condensation (Gottwald & Garner, 1975; Kruschke, 1993; Posner,

This article was published Online First September 22, 2016.

Michael L. Kalish, Department of Psychology, Syracuse University; Ben R. Newell, School of Psychology, University of New South Wales; John C. Dunn, School of Psychology, University of Adelaide.

The support of the Australian Research Council (Grant DP0877510 awarded to the three authors) and the National Science Foundation (Award 125695 to Michael L. Kalish) is gratefully acknowledged. We thank Helen Paton for assistance with data collection.

Correspondence concerning this article should be addressed to Michael L. Kalish, Department of Psychology, Syracuse University, Syracuse, NY 13244. E-mail: mlkalish@syr.edu

1964). *Filtering* tasks are those where it is necessary (or at least, helpful) to filter out one dimension in order to classify items correctly (Posner, 1964, called these *gating* tasks), and *condensation* tasks are those where both dimensions must be evaluated and the results condensed into one judgment. Drawing on the COVIS model of category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby et al., 2011), DeCaro et al. predicted that higher WMC would be beneficial for learning filtering tasks but detrimental for learning condensation tasks.

This rather counterintuitive prediction falls out of the COVIS model because of a secondary assumption of the multiple systems theory. Ashby et al. (2011) hypothesize that the two category-learning systems compete, such that only one can provide a response for the learner to emit at any given point in time. At the outset of learning a new task the competition is biased in favor of the rule-based system, so that early responses are due to its action. If this system is successful in learning the categorization rule, then its responses continue to be used to classify the stimuli. If, however, it receives consistently negative feedback (i.e., successive rules are tested but none are consistently correct) then control over behavior eventually switches to the procedural system which has been learning in the background the whole time. Working memory capacity modulates the initial bias toward the rule-based system; people high in WMC have a stronger bias to use their rule-based system, in part because that system works well for them due to its dependence on WMC. Thus, DeCaro et al. (2008) proposed those participants with higher WMC are likely to persist in the testing of (ultimately futile) rules in the condensation tasks, thus taking longer to relinquish control to the more task-appropriate procedural system and so actually slowing learning relative to participants with lower WMC. DeCaro et al. indeed found that participants with higher WMC took fewer trials than those with lower WMC to reach a learning criterion in a filtering task but more trials in a condensation task.

DeCaro et al.'s (2008) counterintuitive finding was quickly revised, however. Tharp and Pickering (2009) showed in a reanalysis, using a less stringent criterion for data inclusion, that higher WMC led to quicker learning in both tasks. DeCaro, Carlson, Thomas, and Beilock (2009) replicated this result, and also showed that low WMC individuals perseverated with suboptimal rule-based strategies in just the manner originally predicted for high WMC individuals.

Three additional studies have directly questioned COVIS's predictions about the role of WMC in mediating the relative influence of different category learning systems. Craig and Lewandowsky (2012); Lewandowsky (2011), and Lewandowsky, Yang, Newell, and Kalish (2012) used a more comprehensive measure of WMC (Lewandowsky, Oberauer, Yang, & Ecker, 2010), a larger set of category learning tasks, and a repeated-measures methodology. Lewandowsky (2011) trained every participant, over the course of multiple sessions, on six different tasks. These tasks were versions of the six types of categorical distinctions introduced by Shepard, Hovland, and Jenkins (1961), requiring classification of eight stimuli composed of three binary dimensions into two equal-sized categories. While some of these types of categories are captured by easily verbalizable rules, others essentially require that all eight items be memorized. Craig and Lewandowsky (2012) had each of their participants learn two categorization tasks where stimuli were comprised of four binary dimensions. In one of their tasks (the 5–4 task from Medin & Schaffer, 1978) 9 of the 16 possible items are used during training, and there is no simple classification rule. In the other of their tasks (from Medin, Altom, Edelson, & Freko, 1982) 8 of the 16 items are used during training, and they can be classified by applying an exclusive-or rule on two of the four dimensions. Finally, Lewandowsky et al. (2012) trained every participant on sets of six (Experiment 1) or four (Experiment 2) condensation and filtering tasks that varied systematically in difficulty, and all of which involved large numbers of stimuli that varied along two continuous dimensions. In all of these studies people's performance during learning was found to be consistent with a homogeneous beneficial influence of WMC on category learning regardless of the task details.

Despite this evidence to the contrary, multiple systems theorists have maintained the view that category-learning tasks are differentially sensitive to WMC, citing DeCaro et al. (2008) as an example (e.g., Grimm & Maddox, 2013; Maddox & Chandrasekaran, 2014; Miles & Minda, 2011; Rolison, Evans, Dennis, & Walsh, 2012). One possible reason for this might be that the categories used in the studies mentioned above all differ in apparently minor but potentially important ways from the kind of categories often associated with the procedural learning system.

All existing experiments that demonstrate a homogeneous influence of WMC across tasks have used categorical distinctions that are potentially error-free. Lewandowsky's (2011) tasks used only eight stimuli, while Craig and Lewandowsky's (2012) experiments used eight or nine items. With a small number of stimuli, absolute identification might bypass either a rule-based or implicit system, meaning that homogeneous working memory effects in such tasks may not be due to category learning at all. However, Lewandowsky et al.'s (2012) tasks used a large number of stimuli, preventing memorization. Nonetheless, all of their categories were perfectly linearly separable. With perfect separability, participants might have been encouraged to persist in searching for a classification rule past the point at which they would normally shift to an implicit strategy. Thus, the homogeneous effect of working memory might be due to a continuous reliance on a rule-based learning system, as COVIS would claim. To avoid these potential confounds, we chose to adopt two existing pairs of filtering/condensation tasks from within the COVIS tradition (Maddox, Ashby, & Bohil, 2003; Zeithamova & Maddox, 2006), which are taken to rely differentially on the rule-based and procedural systems. In all four of these tasks memorization is essentially impossible and there is no deterministic response strategy that will produce 100% correct responses.

While these tasks are chosen from the existing literature as diagnostic of multiple-systems, their utility depends on the relationship between the categorization strategy people learn to use, which is presumed to differ between filtering versus condensation tasks, and how well people do, which is measured in the proportion of correct responses. This relationship is, however, far from clear; people may try to use the correct strategy but do so poorly, or they may use a suboptimal strategy but do so very well (as in DeCaro et al., 2009). For this reason, it is often necessary to find another measure of categorization strategy, in the form of explicit models that predict the probability with which a participant will classify each item (given its features) as a member of each category. These models are a proxy for cognitive computational modeling, but they have the advantages of being more tractable and less theoretically contentious.

Model comparison is, however, a very difficult exercise even when the models are simple statistical descriptions of behavior. Researchers choose category structures for use in an experiment in the hope that they will be particularly diagnostic of the participants' response strategies. When selecting designs using large numbers of stimuli, Lewandowsky et al. (2012) chose not to use the structures we use here precisely because these are particularly challenging in this regard; different response models make quite similar predictions for these structures. Moreover, when the models in question are not strictly nested, it is difficult to compare them using penalized maximum likelihood as there is no guarantee that the parameters that distinguish the models all provide equal flexibility. Thus, the use of different penalized likelihood measures (such as the Akaike information criterion [AIC] or the Bayesian information criterion [BIC]) may be misleading.

To address these difficulties, we take a number of steps to clarify the influence of working memory capacity on learning. To begin, we measure WMC with the four-part battery of Lewandowsky et al. (2010). Then, we use a two-part approach to our analysis of WMC and learning in condensation and filtering tasks. The first part consists of two analyses of WMC and categorization performance per se: a state-trace analysis (Bamber, 1979) that directly aims to detect a differential influence of WMC on condensation and filtering performance over the course of study, and a Bayesian correlational analysis of WMC and performance.

State-trace analysis (STA) is a general method for determining whether one or more independent latent variables must be postulated to account for the effects of two or more independent variables on two or more outcomes. In the current context, these variables are the time course of learning (early vs. late trials) and, if dual system accounts are correct, the WMC of participants. The two outcomes are performance in the condensation and filtering tasks. STA makes minimal assumptions about the relationship among the independent, latent, and dependent variables and its logic supersedes the dissociation logic sometimes used to characterize that relationship (Dunn, Kalish, & Newell, 2014). However, STA does require that the independent variables be discrete, as it treats them only as nominal measures. Thus, while STA makes fewer assumptions about the distribution of accuracy and WMC scores than does the correlational analysis, it does require discretization of the WMC scores. These two techniques may thus be viewed as complementing each other; differences in the direction of correlation between WMC and accuracy across the two dependent variables would be evidence of a differential influence of WMC on the two tasks, as would a significant deviation from monotonicity (see the Results section, below) in the state trace.

The second part of our analysis pursues a Bayesian model-comparison approach. This allows us to quantify the amount of evidence in favor of particular response models throughout the experiment. Generally, we wish to know whether people with higher WMC have a greater tendency to use rule-like strategies early in training, and if they tend to persist in using such strategies longer than do people with lower WMC. This question is difficult to answer because we must infer strategy use from the goodness of fit of particular models. Thus, the question we can actually answer is whether or not WMC is associated with a higher posterior probability of one or another response model. Our direct estimation approach represents a further advance from the approximate inference approach developed by Donkin, Newell, Kalish, Dunn, and Nosofsky (2015), which in turn is an advance of the practice from merely reporting the best-fitting model or the maximum-likelihood model parameters.

To summarize, in this study we are looking for any differential influence of WMC on category learning across two tasks: condensation and filtering. We use two different pairs of category learning tasks each of which is taken to be diagnostic by multiple-systems researchers: Maddox et al. (2003) for the pair used in Experiment 1, and Zeithamova and Maddox (2006), for the pair used in Experiment 2. If there is the kind of differential influence of WMC on learning these tasks that dual-system theorists propose then we should see correlations between WMC and performance with different signs. If the differential influence is more modest, we should still see a significant departure from a monotonic state trace. Finally, even if differential influence is such that it does not produce a difference in accuracy we should still see that posterior model probabilities are related to WMC. Because the two task pairs differ only with respect to the coordinates used to generate the filtering and condensation categories, differences between them are not of theoretical interest. Therefore, we report both pairs of tasks (which were conducted as successive experiments) together as a single study.

# Method

## Participants

One hundred and sixty undergraduate students from the University of New South Wales participated in return for course credit ($M_{AGE}$ = 19.9, $SD$ = 3.07; 64% female). Eighty participants completed the Experiment 1 pair of the condensation/filtering tasks, while a different 80 participants completed the Experiment 2 pair of tasks.

## Procedure

**Category learning task.** Participants completed each category learning task at individual computers on which stimuli were presented one at a time, in a different random order for each participant. Participants were instructed to place each stimulus into either Category A or Category B by clicking on an appropriate on-screen button with a mouse. Feedback was immediately provided on screen by the word *correct* or *incorrect* which remained below the stimulus for 750 ms. Participants completed three blocks of 80 trials for each task in which the 40 *A* and *B* stimuli were shown once each. Participants were allowed self-timed breaks between each block. Each participant completed one condensation and one filtering task. There was a break between the categorization tasks in which they were told that the new task would involve a new category structure. The order of filtering and condensation tasks was counterbalanced.

The categorization stimuli were Gabor patches, 200 × 200 pixels in size, with varying spatial frequency and spatial orientation generated by sampling randomly from the bivariate normal distributions shown in Table 1. The coordinates used to generate the Experiments 1 and 2 condensation and filtering stimuli were taken from the relevant conditions of Maddox et al. (2003, Exper-

Table 1
*Category Distribution Parameters for the Filtering and Condensation Category Structures Used in Experiments 1 and 2*

| Category structure | $\mu_{frequency}$ | $\mu_{orientation}$ | $\sigma^2_{frequency}$ | $\sigma^2_{orientation}$ | $Cov_{freq\text{-}orien}$ |
|---|---|---|---|---|---|
| | | *Experiment 1* | | | |
| Filtering | | | | | |
| Category A | 285 | 125 | 75 | 9,000 | 0 |
| Category B | 315 | 125 | 75 | 9,000 | 0 |
| Condensation | | | | | |
| Category A | 272 | 153 | 4,538 | 4,538 | 4,463 |
| Category B | 327 | 97 | 4,538 | 4,538 | 4,463 |
| | | *Experiment 2* | | | |
| Filtering | | | | | |
| Category A | 280 | 125 | 75 | 9,000 | 0 |
| Category B | 320 | 125 | 75 | 9,000 | 0 |
| Condensation | | | | | |
| Category A | 268 | 157 | 4,538 | 4,538 | 4,351 |
| Category B | 332 | 93 | 4,538 | 4,538 | 4,351 |

*Note.* The 80 stimuli for the condensation categories were obtained by rotating the filtering stimuli clockwise by 45° around the center of the spatial-frequency/spatial-orientation space and then shifting the spatial frequency and spatial orientation to achieve an appropriate level of discriminability ($d'$). In Experiment 1 the approximate $d'$ was 9 for the condensation task and 3.5 for the filtering task. In Experiment 2 the approximate $d'$ was 6.7 for the condensation task and 4.3 for the filtering task. (Higher levels of $d'$ indicate more discriminable stimuli.)

iment 2) and Zeithamova and Maddox (2006, Experiment 1), respectively. In Experiment 1 the condensation task had a higher discriminability, $d'$ (which refers to the average distance from stimuli to the category boundary; see Table 1) than did the filtering task and performance on condensation was better than on filtering (see the Results section). In Experiment 2 the distributions had more nearly equivalent $d'$ values (see Table 1) and also equivalent levels of accuracy on the two tasks (see the Results section). Both condensation designs allow approximately 70% long-term accuracy with an optimal one-dimensional rule, but approximately 90% accuracy with an optimal two-dimensional boundary. The stimuli were produced using MATLAB (Mathworks, Natick, MA) routines from the Psychophysics Toolbox (Brainard, 1997), and were presented centered on the computer screen.

**Working Memory Test Battery.** Participants individually completed four computerized working memory tests: an operation span (OS), a sentence span (SS), a memory updating (MU), and a short-term spatial memory test (SSTM) (for details, see Lewandowsky et al., 2010). All tests were programmed and delivered via MATLAB. The order of the four working memory (WM) tests was fixed (MU-OS-SS-SSTM) but the order of completing the two category learning tasks and the WM battery was counterbalanced across three experimental sessions, each of which lasted approximately 1 hr.

# Results

## Working Memory Tasks

Summary statistics for the four WM tasks are displayed in Table 2. These values are comparable across the experiments and with previous studies using these tests (cf. Lewandowsky et al., 2010). Lewandowsky et al. reported that this test battery has very high validity, both internal and external. Our participants' performances share similar central tendencies, ranges and variances with Lewandowsky et al.'s, leading us to believe that the psychometric properties described in the development of these measures extend to our observations. There was no credible difference in the distributions of the subscales between Experiments.

In line with the development of these measures, and consistent with the subscale correlations shown in Table 2, we assume that they represent aspects of a single underlying factor that we identify as WMC. We used confirmatory factor analysis to find the loadings of the working memory scores on this factor, with the results shown in Table 2. The factor accounts for approximately 61% of the variance in the scores. Each participant was assigned a "true" WMC score using these loadings.

## Accuracy

Accuracy on each of the three blocks for each of the two tasks for each of the two experiments is presented in Figure 1. In Experiment 1, where the $d'$ of the filtering task is low, it is more difficult than the condensation task. In Experiment 2 the two tasks are of more equivalent difficulty. We used two different analytic techniques to look for associations between WMC scores and category learning accuracy: (1) an investigation of differences in linear correlations between WMC and accuracy by task and (2) a general search for differential influence of WMC using state-trace analysis.

## Correlation Analysis

As an initial analysis, we computed the correlations between each WMC measure and percentage correct accuracy for each of the three

Table 2
*Descriptive Statistics of the Four Working Memory Tests in Experiments 1 and 2 (N = 80), With the Overall Subscale Correlations and the First Principal Component*

| Measure | MU | OS | SS | SSTM |
|---|---|---|---|---|
| | | *Experiment 1* | | |
| Mean | .58 | .71 | .64 | .86 |
| Standard deviation | .19 | .15 | .20 | .06 |
| Minimum | .20 | .27 | .23 | .69 |
| Maximum | .97 | .99 | .95 | .98 |
| Skewness | −.68 | .78 | −.88 | .20 |
| Kurtosis | .22 | −.86 | −.48 | −.61 |
| | | *Experiment 2* | | |
| Mean | .55 | .62 | .62 | .86 |
| Standard deviation | .19 | .22 | .21 | .07 |
| Minimum | .15 | .08 | .06 | .53 |
| Maximum | 1.00 | .97 | .98 | .99 |
| Skewness | −.64 | .00 | −.19 | 4.89 |
| Kurtosis | .09 | −1.01 | −.55 | −1.64 |
| Standardized loadings | .70 | .65 | .55 | .53 |
| Correlations | | | | |
| MU | | .463 | .326 | .415 |
| OS | | | .391 | .234 |
| SS | | | | .332 |

*Note.* MU = memory updating; OS = operation span; SS = sentence span; SSTM = spatial short-term memory.

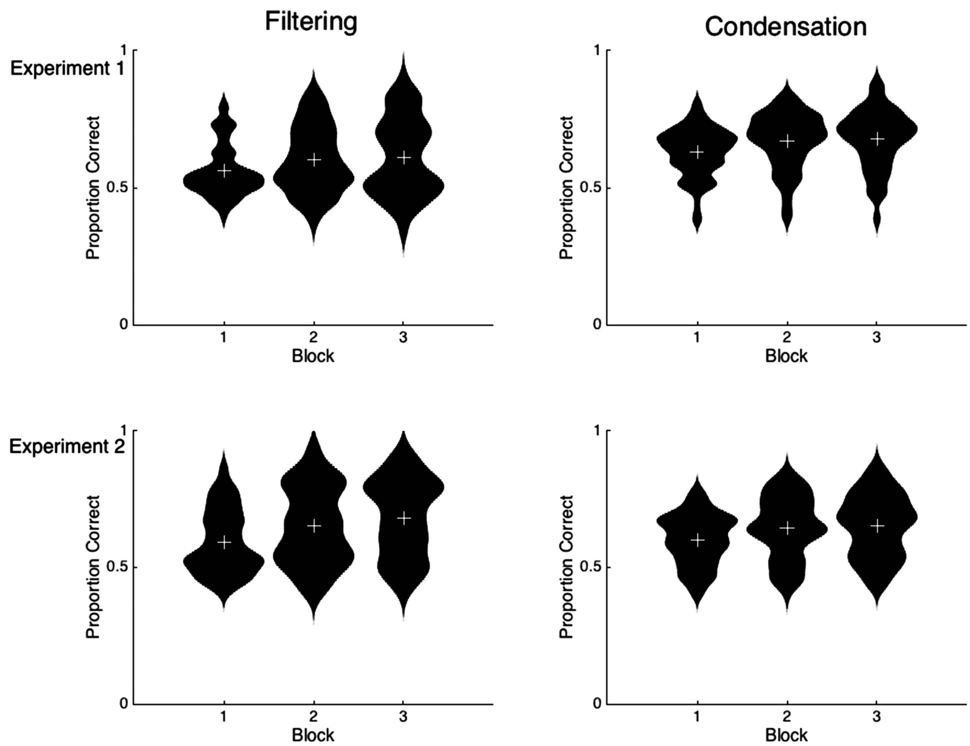


*Figure 1.* Accuracy of categorization for each of the three blocks of the two experiment's versions of the condensation and filtering tasks. Violin plots show the mean densities (white crosses) and smoothed kernel densities (black regions). See the online article for the color version of this figure.



blocks of trials and for condensation and filtering separately, combining both experiments. We used a Bayesian estimation approach to compute the posterior distribution of each of these six correlations independently.[1] The results are shown in Table 3. For all three blocks of the filtering task the credible regions included zero. For all three blocks of the condensation task the credible regions were entirely greater than zero. However, the high-density intervals are quite broad, indicating that the precision of the estimate is rather low. Only for the final block of the condensation condition does the 95% high-density interval entirely exclude a reasonable region of practical equivalence of (−.10, .10).[2] However, despite the low precision of the estimates, it is clear that there is no reliable negative correlation at any stage for the condensation task. Thus, the correlation analysis suggests a modest positive relationship between WMC and average performance overall (the median posterior $r = .16$), which is substantially larger for the later blocks of the condensation condition but which is too small to be considered evidence for a relationship in the filtering condition.

Table 3
*Estimated Correlations Between Working Memory Capacity and Categorization Accuracy at Each Block of Training, Combined Across Experiments (N = 160)*

| Condition | Block | Median of posterior correlation | Lower bound of HDI | Upper bound of HDI |
|---|---|---|---|---|
| Filtering | 1 | −.099 | −.258 | .054 |
| | 2 | .020 | −.146 | .167 |
| | 3 | .095 | −.062 | .249 |
| Condensation | 1 | .212 | .057 | .363 |
| | 2 | .239 | .088 | .387 |
| | 3 | .254 | .106 | .401 |

*Note.* HDI = high-density interval.

## State-Trace Analysis

The hypothesis under investigation is that WMC is positively related to performance on the filtering task but negatively related to performance on the condensation task, at least initially. The previous correlation analysis depends on an assumption of linearity in the relationship between the latent mastery of the category structures and

[1] We used the code provided by Bååth (2014), which specifies a model as follows. The data (percentage correct, working memory capacity) are taken to be distributed as a bivariate normal. The prior on the means are normal, with a mean of the sample mean and broad variance. The prior on the covariance matrix is broken into on-diagonal and off-diagonal elements. The priors for off-diagonal elements are broach uniform distributions ranging from 1/1,000 to 1,000 times the observed variance. The priors on the parameter of interest, the on-diagonal correlation coefficient, is uniform over (−1,1). All chains were well mixed and were nonautocorrelated.

[2] We also performed a Bayes factor analysis using JASP (JASP Team, 2016), which approximates the posterior analytically. Only Blocks 2 and 3 of the condensation condition provided evidence against the null of no correlation ($BF_{10} > 10$). However, it is also the case that no block of data provided evidence in favor of a negative correlation; the highest $BF_{10}$ was for Block 1 of the filtering condition, which was only 0.40.

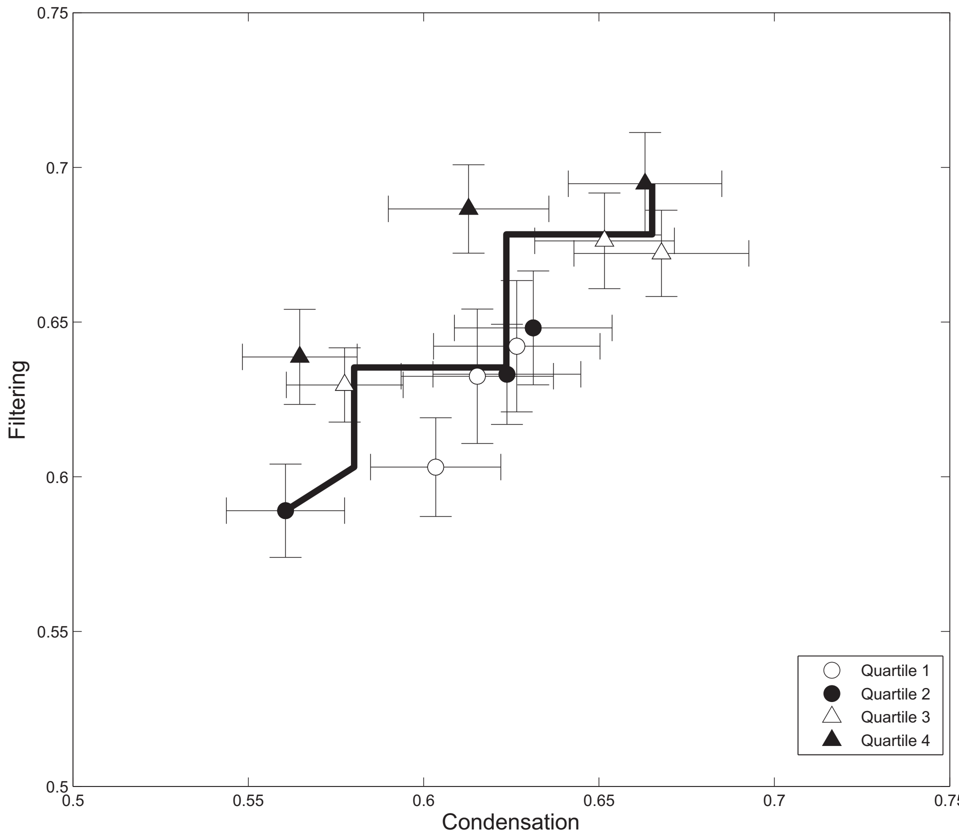


*Figure 2.* State-trace plots of the performance on the condensation and filtering tasks of participants divided into quartiles by working memory capacity (WMC). (A) Results from all participants. The data are shown for the first (lowest, Quartile 1) through fourth (highest, Quartile 4) quartiles of WMC. Error bars are the sample standard errors of the mean. The best fitting one-dimensional state trace, found by coupled monotonic regression, is shown as the solid line. Statistical analysis reveals the deviation from this regression to be nonsignificant.



the observed performance on the categorization tasks. A stronger test of the hypothesis that WMC differentially affects filtering and condensation is provided by state-trace analysis (STA), as it does not depend on this strong linearity assumption (Dunn, Newell, & Kalish, 2012; Dunn, Kalish, & Newell, 2014). In order to perform STA, all independent variables need to be discrete. We therefore quantized WMC by placing participants into four quartiles of 40 participants each. Performance of these groups on the two tasks is shown in the state-trace plots in Figure 2.

The signature of the differential influence of WMC on learning filtering categories is for the data in Figure 1 to not lie on a single monotonic curve (Dunn et al., 2014). We found the best-fitting monotonic curve through the data in Figure 1 using the coupled monotonic regression method developed by Kalish, Dunn, Burdakov, and Sysoev (2016). Because Bayesian estimation methods have not yet been developed for this design, we used Kalish, et al.'s parametric resampling bootstrap procedure to provide an empirical $p$ value to evaluate the null hypothesis that monotonic curve is an adequate description of the data. The observed value for the one-dimensional state-trace was $p > 0.29$, indicating that we cannot reject the null hypothesis of a uniform influence of WMC on the two tasks.[3] Any discretization of a continuous variable has the potential to produce misleading results. We chose four levels rather than a median split (two levels) to ensure that at least some of the resulting quantiles are genuinely different, and 4 levels rather than, for example, 10 levels to ensure that there were enough measurements within each level to allow good estimation of the dependent variables. However, as a check we computed a $p$ value for quantizations ranging from 2 to 10 groups. In each case the result was nonsignificant.

Despite these results, one might still argue that the nature of the involvement of WMC in learning, and its differential influence on the condensation and filtering tasks, remains somewhat ambiguous. COVIS posits that the main role of WMC in learning is to bias participants toward using a rule-based strategy, regardless of the task at hand. That is, people with higher WMC should be inclined to persist in using a rule-based strategy, while people with lower WMC should not—they should either abandon the rule-based strategy in the condensation task or never attempt it in the first

[3] We also considered the standard analysis of variance approach to dissociations with these data, which found only a marginal interaction among WMC, task, and block. This interaction is not diagnostic of whether there was, or was not, differential influence, however.

place. Categorization accuracy is not uniquely related to categorization strategy. People with higher WMC might perform just as well by using their rule-based strategy (recall that both condensation designs allow approximately 70% long-term accuracy with an optimal one-dimensional rule) as people with low WMC might by using a similarity-based strategy (where errors might be frequent early in learning). We therefore turn to a formal analysis of response strategies per se, rather than accuracy, to resolve the role of WMC in these tasks.

## Response Strategy Analysis

In this analysis, we fit sets of response-surface models to the data from each participant so as to identify their categorization strategy. The models are described formally in the Appendix and depicted in Figure A1; here we describe them informally. A response surface model is one which simply predicts the probability a participant will apply label *A* to any given stimulus, based solely on its values of orientation and spatial frequency. We chose as our least restricted model one that assumes that both *A* and *B* items are normally distributed with equal variance, with the means of these category distributions able to vary in location on both the spatial frequency and orientation dimensions. This model, which is a naïve Bayes classifier, effectively places a linear boundary of arbitrary orientation at an arbitrary location in the two- dimensional space, with a gradient, or gain, determined by its variance parameter. Because distribution locations may vary in two dimensions, we call this a two-dimensional (2D) model. It is a moderately complex model, having three free parameters (two for location, one for variance), and was fit to each block of the experiment separately.

We considered two kinds of models with fewer than three parameters. The first is our null model, in which the participant is held to simply guess *A* at a constant, freely estimated, rate. The second consists of naïve Bayes classifiers that only use one of the two stimulus features, which we call 1D models. In the filtering conditions, where only spatial frequency matters, the one-dimensional (or 1D) spatial frequency classifier would be expected to describe an ideal observer. We fit both possible one-dimensional models, which have two parameters each, to each block of data. Each of these models effectively places a decision boundary at some value of either spatial frequency or orientation, with a gradient determined by the common variance of the *A* and *B* distributions.

Finally, we considered two additional models that had three free parameters, but which are more constrained than the first model. Response surfaces with multiple boundaries, each aligned with a stimulus dimension, have been considered as plausible rules in categorization, so we generated two such bilinear models. Basically, each model uses two one-dimensional naïve Bayes classifiers and then applies either a conjunctive or a disjunctive decision rule to their output. Essentially, each model carves the space into (not necessarily equal area) quadrants; one model identifies three of these quadrants as *A* and one model identifies three of these quadrants as *B*. These 2D models, like the unrestricted classifier, use all the information in the stimulus, unlike the two 1D models.

Overall, then, there are six models to be compared. Three 2D models depend on both stimulus features: the unconstrained linear two-dimensional classifier and the two constrained bilinear classifiers. Two 1D models depend on only one stimulus feature: These are the one-dimensional linear classifiers. Lastly, there is the random guessing model that can fit otherwise unclassifiable blocks of responses.

We pursued a Bayesian estimation approach to model comparison (Kruschke, 2014). This approach avoids having to apply an arbitrary correction for the number of parameters in each model, as frequentist likelihood-based approaches must do.[4] The details of our procedure are provided in the Appendix, but essentially it produces a direct estimate of the posterior probability of each model given the data. Figure 3 shows the posterior probability of each of the six models, represented in one of four shades of gray, for each participant in each of the three blocks of both the filtering and condensation tasks. The participants are arranged in increasing order of WMC in each graph to allow visual comparison of the patterns of model fit over blocks and conditions. Because our concern is with the posterior probabilities of model classes, the two 1D models are both shaded dark gray, while the conjunctive and disjunctive models are both shaded light gray. The 2D linear model is white and the random responding model is black.

The first thing to notice is that there are clear changes in the distributions of posterior probability across blocks for both tasks. In the first block of filtering training (upper row of panels), where the "correct" model would be the 1D discrimination on spatial frequency, the 2D models have most of the mass. This is largely due to these models' ability to capture more types of variability in the responses, which are structured enough to reject the null, random-guessing, model but too noisy to favor the restricted 1D model. As the experiment continues, the participants gradually learn to give the correct responses, and the posteriors converge on the spatial frequency 1D model. Inspection of Figure 3 suggests that there is, however, no relationship between the participants' WMC and the model posteriors either within or between blocks. We computed the correlation between WMC and model posterior within each block. The only marginally credible correlations were observed in the third block, with a small positive relationship between WMC and the posteriors of the 2D linear and the posteriors of the 2D conjunctive model ($r = .234$ and $r = .232$, respectively).[5]

The condensation task (lower row of panels, Figure 3) shows similar results, but with the direction of change reversed. Early in the experiment participants are generally consistently labeling the items by either their spatial frequency or orientation. By the end of the experiment, however, participants were more likely to produce responses consistent with one of the 2D models. The conjunctive classifier and the bivariate linear classifier appear to fare about equally well, which is not surprising given that the structure of these categories does little to discriminate these two classification schemes (a point raised in the design of Lewandowsky et al.'s, 2012, experiments). Even more than with the filtering task, there is a lack of any consistent change in response model posterior distributions across the

[4] We also pursued a maximum likelihood approach, comparing these models using both AIC and BIC. The relative goodness of fit of the various models did indeed depend on which penalty was imposed, but not in any obviously systematic way.

[5] We also correlated WMC with the interblock differences in posterior probability for each model (e.g., the posterior of the 1D models in Block 1 vs. Block 2 of the condensation condition), and with the intercondition differences (e.g., the posterior of the 2D linear model in Block 1 of the filtering vs. Block 1 of the condensation condition) and found no credible relationships.

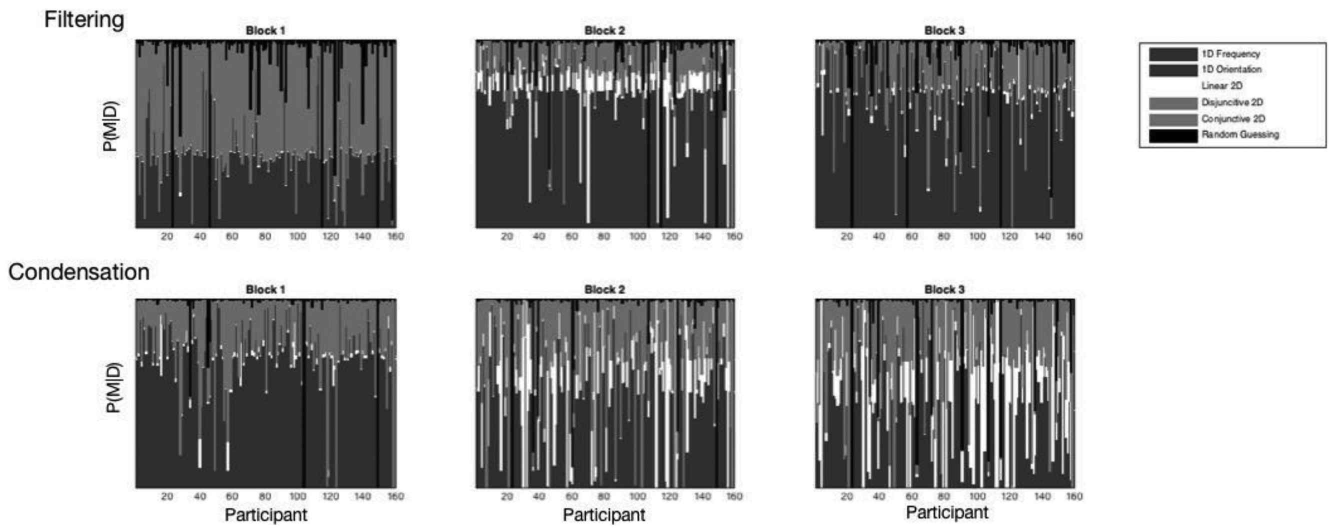


*Figure 3.* The posterior probabilities of each of the response models given the data, $p(M_i|D)$, for each model *i*, for each participant at each of the three blocks in each of the two conditions. The top row is the filtering task, and the bottom row is the condensation task. Participants are ordered by their working memory capacity working memory capacity composite scores. The two one-dimensional (1D) models are colored the same, as are the two two-dimensional (2D) bilinear models (see the text), to allow easier interpretation of the figure.

ordinate (the largest $|r|$ was less than 0.16), reflecting a lack of any differential bias or selective influence of WMC on strategy choice.

## Discussion

Working memory is conceptualized as a fundamental cognitive resource, and categorization is a fundamental cognitive capacity. It would not be surprising if working memory capacity was positively related to a person's ability to learn to remember new categorical distinctions. Categorization, however, is a complex capacity. There are likely many cognitive mechanisms involved in supporting it, and many strategies at play in the way it is exercised. It would not be surprising if working memory capacity was related to category learning only relatively weakly, or more strongly in some situations than others. What would be surprising, from a pretheoretical perspective, would be if having a larger working memory capacity actually resulted in worse performance on a category learning task. This, of course, is exactly the claim made by COVIS (Ashby et al., 2011) when the task in question is one which requires a person to surrender explicit control of their learning and allow it to happen implicitly.

According to the logic of COVIS, one data pattern that would support this view of two independent learning systems is as follows: (1) significant positive correlations between accuracy and WMC on the filtering task; and (2) significant negative correlations between accuracy and WMC on the condensation task, especially early in training. Table 3 shows that this pattern was not found. We found that the only credible correlations were for the condensation task, where higher WMC was clearly associated with better performance.

The differential dependence of category learning systems on WMC makes two additional parameter-free predictions. The first is that performance on condensation and filtering tasks should produce a two-dimensional state trace when high- versus low-WMC participants are tested across blocks. As Figure 2 shows, however, the data are consistent with a one-dimensional state trace. Our frequentist test cannot support the null here, but the observed state trace is not significantly two dimensional. Second, the strategy a participant chooses to use, and thus the response surface model that best describes their data, should differ with WMC. Our analysis, summarized in Figure 3, is designed to reveal any such bias as a systematic change in the model posterior distributions across the ordinate. No such trend is in evidence demonstrating that there is no systematic preference for 1D models or the 2D conjunctive model by high-WMC participants at any stage of learning the condensation task.

The exchange regarding the involvement of WMC in category learning (DeCaro et al., 2008, 2009; Tharp & Pickering, 2009) strikes at the heart of the multiple- versus single-system debate. Our two experiments find no support for a system that operates independently of working memory in the domain of perceptual category learning. Speaking generally, it appears that working memory tends to aid performance across category structures in general.

Interestingly, this facilitative effect was observed in the putatively "implicit" condensation task and not in the "explicit" filtering task. This might be because variability in filtering performance is related more to perceptual factors than to memorial ones (e.g., the ability to tell what value a stimulus has on a given dimension, rather than to remember what value the criterion is). However, such a result would also be predicted if one assumes that participants approach both condensation and filtering tasks in the same manner, that is, by trying—explicitly—to find the (perhaps quite complex) boundary or representation that delivers accurate categorization. When the boundary is not easily verbalizable, as in the condensation task, one might expect that trying to find and maintain the correct response strategy requires working memory resources and thus leads to a benefit for those with higher WMC. This is what we found, with WMC associated with better performance in later stages of condensation learning but not associated with any difference in the probability of using a complex response strategy. Thus, the kind of strategy people use during learning the condensation task does not differ with WMC, but their command of that strategy does, meaning that they perform better but are not classified differently. In the filtering task, by contrast, there is no overall benefit of high WMC in terms of performance. This may be because the primary difficulty in the filtering task may be unrelated to WMC, lying instead in the problem of accurately estimating the spatial frequency of the to-be-classified stimulus. The fact that performance is not unambiguously identified with the 1D classifier may be due to the difficulty of this decision, or to a general dissatisfaction with this difficult strategy leading to continued exploration of other strategies.

## Single Versus Dual Systems

Just because our data are not consistent with the predictions of a particular multiple systems view does not, of course, mean that they are necessarily consistent with any particular single-systems theory. Single-system theories typically do not offer the sort of structural predictions that a theory like COVIS makes. We suggest that two lines of analysis would be required to shed more light on the role of WMC in condensation and filtering tasks. The first is the identification of cases when it is possible to observe the differential influence of any relevant factor (such as WMC) on any relevant outcome (such as accuracy in category-learning tasks). STA tests just this kind of generic distinction, and showed, in the current case, that there was an absence of evidence for multiple latent process governing learning of condensation and filtering tasks. The value of pursuing this approach across a range of tasks and conditions would be to develop a theory-neutral catalog of data that characterize how category learning is articulated. In essence, this would suggest where the “joints” were in the “body” of this capacity, allowing its successful dissection. The second line of analysis would be to develop a structural model of our tasks using a quantitative single-system model such as the Generalized Context Model (Nosofsky, 1986) or attention learning covering map model (Kruschke, 1993). While we suspect that this should be feasible, we have not pursued theory-based modeling here because we believe that the effort should be made in the context of a survey of the larger body of results that characterize category learning.

Our measure of WMC was derived from four component scores, but we analyzed the effects of WMC as a unitary construct. We did this, in part, to avoid a “garden of forking paths” problem (Gelman & Loken, 2016). In part, however, we did so because the claim under investigation is whether an “implicit” or “procedural” system makes an independent contribution to category learning. However, the development of a structural model might require closer investigation of these components. An exploratory analysis of the association of category learning performance across both category learning tasks with the SSTM component of WMC found that correlations were credibly positive on all blocks of both tasks with the exception of the first block of the filtering task. The SSTM task requires participants to remember and to reproduce the location of 2–6 dots in a $10 \times 10$ grid. It is plausible that good memory for such spatial coordinates is likely to be useful when trying to learn about categories defined by the width and orientation of sine-wave gratings. None of the other components were individually credibly associated with performance across blocks and conditions.

It is of interest in this context that at least one set of results interpreted as supporting the multiple-systems view (Miles & Minda, 2011) also show that learning a condensation structure is disrupted by a concurrent visual working memory task. The cross-task role of visual working memory in learning has also been accepted in recent interpretations of the COVIS model (Valentin, Maddox, & Ashby, 2014). Thus, while our results are not consistent with a view that there exists an implicit or procedural learning system operating independently, they are consistent with recent findings from other category learning studies. This leads naturally to the question of whether the involvement of visual “working memory,” which is necessarily explicit, in category learning can reasonably be conceptualized as a property of a procedural system that is, by definition, implicit. The methodology we have used in the current study could, of course, be applied to address questions about the influence of visual working memory if such questions were considered to be theoretically relevant.

Our study used two sets of category structures taken from the existing literature. However, we found that these particular structures are only weakly diagnostic of any differences in strategy use. The Bayesian model posteriors were generally ambivalent between 1D and at least one of the 2D models, which made it impossible to rule out either of these classes of strategy for the vast majority of participants. This nondiagnosticity echoes the concerns that led Lewandowsky et al. (2012) to adopt novel category structures for their investigation. Nevertheless, our evaluation of model posterior probabilities represents an advance in analyzing multiple overlapping strategies in category learning experiments. Our modeling results appear to demonstrate the absence of any effect of WMC on categorization strategy early in learning, while possibly suggesting some subtle differences late in learning. Critically, the results provide a full picture of the uncertainty inherent in identifying just how participants are learning the categorization tasks. We suggest that future research into the number and nature of latent factors governing category learning will need to carefully consider their designs, and that historical concerns should not dominate theoretical ones. Methods that attempt to identify learners’ strategies or “systems” via maximum likelihood estimates of the fit of response surface models are prone to overstating the evidence such estimates provide. Donkin et al. (2015) have made a similar point in the context of demonstrating the difficulties inherent in classifying participants’ response strategies in a standard four-category design.

Many other variables “dissociate” performance on condensation and filtering tasks, leading some to argue that there is overwhelming support for multiple-systems views (e.g., Ashby et al., 2011; Valentin et al., 2014). However, our results and other recent findings (Craig & Lewandowsky, 2012; Lewandowsky, 2011; Lewandowsky et al., 2012) suggest that such views should be modified in light of the fact that when WMC contributes to learning it does so positively to learning both condensation and filtering category structures. An understanding of the psychological mechanisms responsible for this positive influence remains elusive. Neither dissociation logic nor STA can count the number of systems involved, but only the minimum number of latent variables, as pointed out by Dunn et al. (2014). Further research will be needed to articulate just how people use their working memory during category learning.

## References

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. http://dx.doi.org/10.1037/0033-295X.105.3.442

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

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. http://dx.doi.org/10.1017/CBO9780511921322.004

Bååth, R. (2014, June). *Bayesian first aid: A package that implements Bayesian alternatives to the classical *.test functions in R.* Paper presented at UseR! 2014: The R User Conference, Los Angeles, CA.

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19,* 137–181.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 65,* 439–464. http://dx.doi.org/10.1080/17470218.2011.608854

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,* 397–403. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.cognition.2007.07.001

Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2015). Identifying strategy use in category learning tasks: A case for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 933–948. http://dx.doi.org/10.1037/xlm0000083

Dunn, J. C., Kalish, M. L., & Newell, B. R. (2014). State-trace analysis can be an appropriate tool for assessing the number of cognitive systems: A reply to Ashby (2014). *Psychonomic Bulletin & Review, 21,* 947–954. http://dx.doi.org/10.3758/s13423-014-0637-y

Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: The limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 840–859. http://dx.doi.org/10.1037/a0027867

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

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist, 102,* 460–465. http://dx.doi.org/10.1511/2014.111.460

Gottwald, R., & Garner, W. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception & Psychophysics, 18,* 26–28. http://dx.doi.org/10.3758/BF03199362

Grimm, L. R., & Maddox, W. T. (2013). Differential impact of relevant and irrelevant dimension primes on rule-based and information-integration category learning. *Acta Psychologica, 144,* 530–537. http://dx.doi.org/10.1016/j.actpsy.2013.09.005

Horsey, R. (2002). The art of chicken sexing. *UCL Working Papers in Linguistics, 14,* 107–117.

JASP Team. (2016). *JASP* (Version 0.7.5.5) [Computer software]. Available from https://jasp-stats.org/

Kalish, M., Dunn, J., Burdakov, O., & Sysoev, O. (2016). A statistical test for the equality of latent orders. *Journal of Mathematical Psychology, 70,* 1–11. http://dx.doi.org/10.1016/j.jmp.2015.10.004

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science, 5,* 3–36. http://dx.doi.org/10.1080/09540099308915683

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R and BUGS* (2nd ed.). Burlington, MA: Academic Press.

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. http://dx.doi.org/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. http://dx.doi.org/10.1037/a0022639

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. http://dx.doi.org/10.3758/BRM.42.2.571

Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 881–904. http://dx.doi.org/10.1037/a0027298

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. http://dx.doi.org/10.1037/0278-7393.29.4.650

Maddox, W. T., & Chandrasekaran, B. (2014). Tests of a dual-system model of speech category learning. *Bilingualism: Language and Cognition, 17,* 709–728. http://dx.doi.org/10.1017/S1366728913000783

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

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

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. http://dx.doi.org/10.1037/a0022309

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/S0079-7421(10)52003-6

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

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

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. H. Ross (Ed.), *The psychology of learning & motivation* (Vol. 54, pp. 167–215). Amsterdam, the Netherlands: Elsevier. http://dx.doi.org/10.1016/B978-0-12-385527-5.00006-1

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. http://dx.doi.org/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. http://dx.doi.org/10.3758/BF03196274

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255.

Palmeri, T. J., & Flanery, M. A. (2002). Memory systems and perceptual categorization. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research theory* (Vol. 41, pp. 141–189). San Diego, CA: Academic Press.

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

Posner, M. I. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review, 71,* 491–504. http://dx.doi.org/10.1037/h0041120

Rolison, J., Evans, J., Dennis, I., & Walsh, C. (2012). Dual-processes in learning and judgment: Evidence from the multiple cue probability learning paradigm. *Organizational Behavior and Human Decision Processes, 118,* 189–202. http://dx.doi.org/10.1016/j.obhdp.2012.03.003

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

Speekenbrink, M., Channon, S., & Shanks, D. R. (2008). Learning strategies in amnesia. *Neuroscience and Biobehavioral Reviews, 32,* 292–310. http://dx.doi.org/10.1016/j.neubiorev.2007.07.005

Stanton, R. D., & Nosofsky, R. M. (2007). Feedback interference and dissociations of classification: Evidence against the multiple-learning-systems hypothesis. *Memory & Cognition, 35,* 1747–1758. http://dx.doi.org/10.3758/BF03193507

Tharp, I. J., & Pickering, A. D. (2009). A note on DeCaro, Thomas, and Beilock (2008). Further data demonstrate complexities in the assessment of information-integration category learning. *Cognition, 111,* 410–415. http://dx.doi.org/10.1016/j.cognition.2008.10.003

Valentin, V. V., Maddox, W. T., & Ashby, F. G. (2014). A computational model of the temporal dynamics of plasticity in procedural learning: Sensitivity to feedback timing. *Frontiers in Psychology, 5,* 643. http://dx.doi.org/10.3389/fpsyg.2014.00643

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. http://dx.doi.org/10.3758/BF03196154

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



# Appendix

## Response Strategy Models

The models used to describe performance are the same as those developed by Lewandowsky et al. (2012) and described in their Appendix. Here we briefly rehearse that description and further explain the hierarchical Bayesian model used to estimate the posterior distributions of the model probabilities and parameters. Prior to modeling, the stimulus values were standardized to fall in the interval (0, 1) on both dimensions.

There are six models of the choice probability $p(A|x)$ all of which hold that $p(A|x) = 1 - p(B|x)$. The five learning models (exclusive of random guessing) hold that the likelihood $p(x|A)$ is given by a bivariate Gaussian distribution with a mean vector, $\boldsymbol{\mu}_A$, and a standard deviation $\sigma$ that is the same for both stimulus dimensions; these models are shown in the five panels of Figure A1. For convenience, and without loss of generality, we set $\boldsymbol{\mu}_A$ equal to the null vector by subtracting the mean of the stimuli given label *A* by the participant from the values of each stimulus; in Figure A1 the circle representing Category A is located at (0, 0) in every panel. The five learning models differed by using different representations of $p(x|B)$, as described below. The sixth model is a biased random guessing model which holds that $p(A) = \pi$, where $\pi$ is a free parameter.

The general linear classifier defines the likelihood, $p(x|B)$, as a bivariate normal, $N(\boldsymbol{\mu}_B, \sigma\mathbf{I})$, and is shown in the leftmost panel of Figure A1. The posterior distribution $p(A|x)$ has an equiprobability boundary that forms a straight line at a particular location and orientation, both determined by $\boldsymbol{\mu}_B$, and with a gradient (i.e., the rate of change of the choice probability) determined by the ratio of $\sigma$ to the magnitude of $\boldsymbol{\mu}_B$. Because $\boldsymbol{\mu}_B$ is free to vary on both stimulus dimensions, we call this a two-dimensional model

When one of the components of $\boldsymbol{\mu}_B$ is set to zero (as in the two upper panels on the right side of Figure A1), the resulting distribution has an equiprobability boundary that forms a straight line parallel to the dimension corresponding to the non-zero component. The location and gradient of the boundary are again determined by $\boldsymbol{\mu}_B$ and $\sigma$. We identify these as one-dimensional models, because prediction depends on only one of the two stimulus dimensions.

We used the same formalism to construct models consistent with two-dimensional conjunctive or disjunctive rules. 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 $\sigma$. Similarly, $p(x|B_2)$ is given by a distribution with mean $(0, B_2)$. The likelihood for the disjunctive model is just $p(x|B) = \min(p(x|B_1), p(x|B_2))$. 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 a function of $\sigma$. These are also 2D models, and are shown in the bottom two panels on the right side of Figure A1.

(*Appendix continues*)

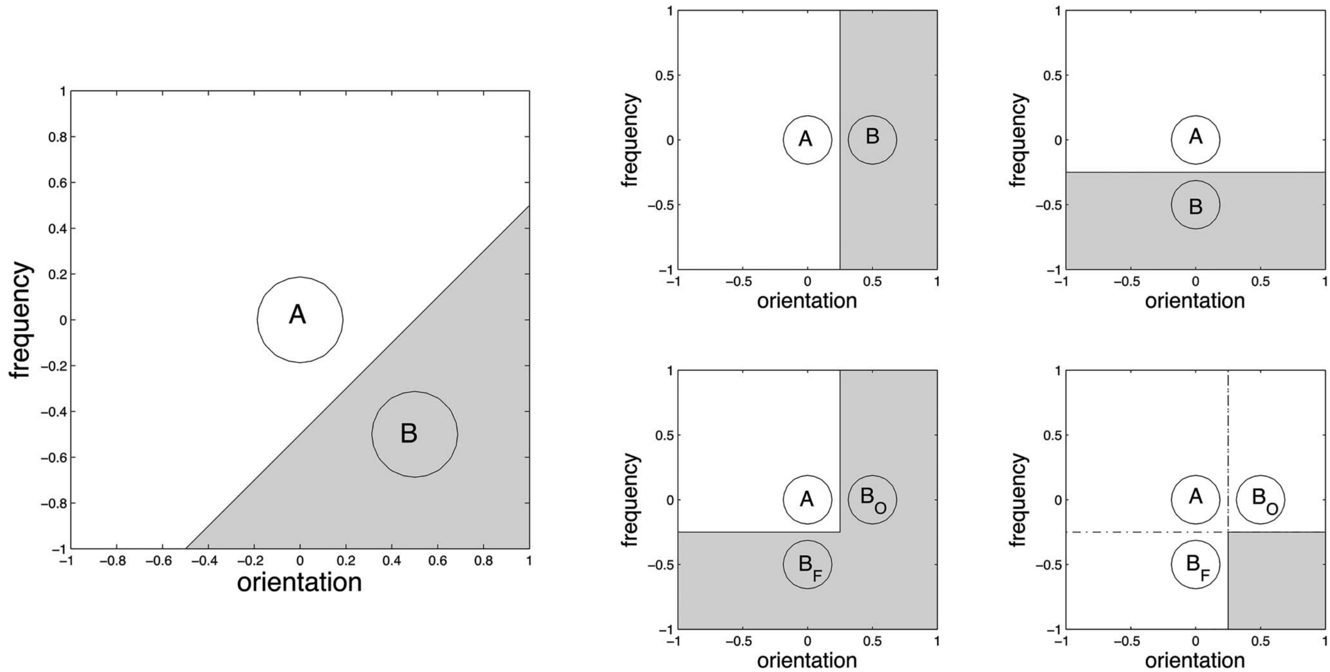


*Figure A1.* The five learning models used in the strategy analysis. The left column shows the two-dimensional model. Categories are assumed to be equal-variance Gaussian distributions, with the mean of Category A set to (0, 0), with the mean of *B* on both dimensions and the shared variance as three free parameters. The circles represent the set of stimuli *X* with equal likelihood p(*X*|Category). The top row shows the two one-dimensional models, each with the mean of *B* set to equal zero on one of the two dimensions. The bottom row shows the bilinear full models, with locations of the B means shown as though they were distributions. Unshaded regions are more likely to be classified as A.

We placed these models in a hierarchy by assuming that each subject is a random multinomial draw from a Dirichlet distribution over models, and that this distribution is different for each of the three training blocks in each of the two tasks. The Dirichlet distribution itself is drawn from a uniform prior over each of the six model types. Each component of $\boldsymbol{\mu}_B$ for each model is given a uniform prior over $(-1, 1)$, $\sigma$ is given a uniform prior over (0.1, 5), and the guessing parameter $\pi$ is given a uniform prior over (0, 1).

To estimate the posterior probability of each model for each participant we used a transdimensional Markov-chain Monte Carlo method (Kruschke, 2014). This method requires computation of pseudo-priors, which were found by estimating the posteriors of each model's parameters for each participant in each block for each task. With these pseudo-priors in hand, we used JAGS to collect four chains of 10,000 samples each with a 1,000 sample burn-in. Chains were well mixed and aggregated for use in estimating the posteriors of all model parameters and hyperparameters. The only parameters of interest are the model probabilities for each block, which are reported in Figure 2.

Received March 30, 2016
Revision received June 24, 2016
Accepted June 28, 2016 ■