Supplemental Material 1

Pigeon Category Learning: Revisiting the Shepard, Hovland, and Jenkins (1961) Tasks

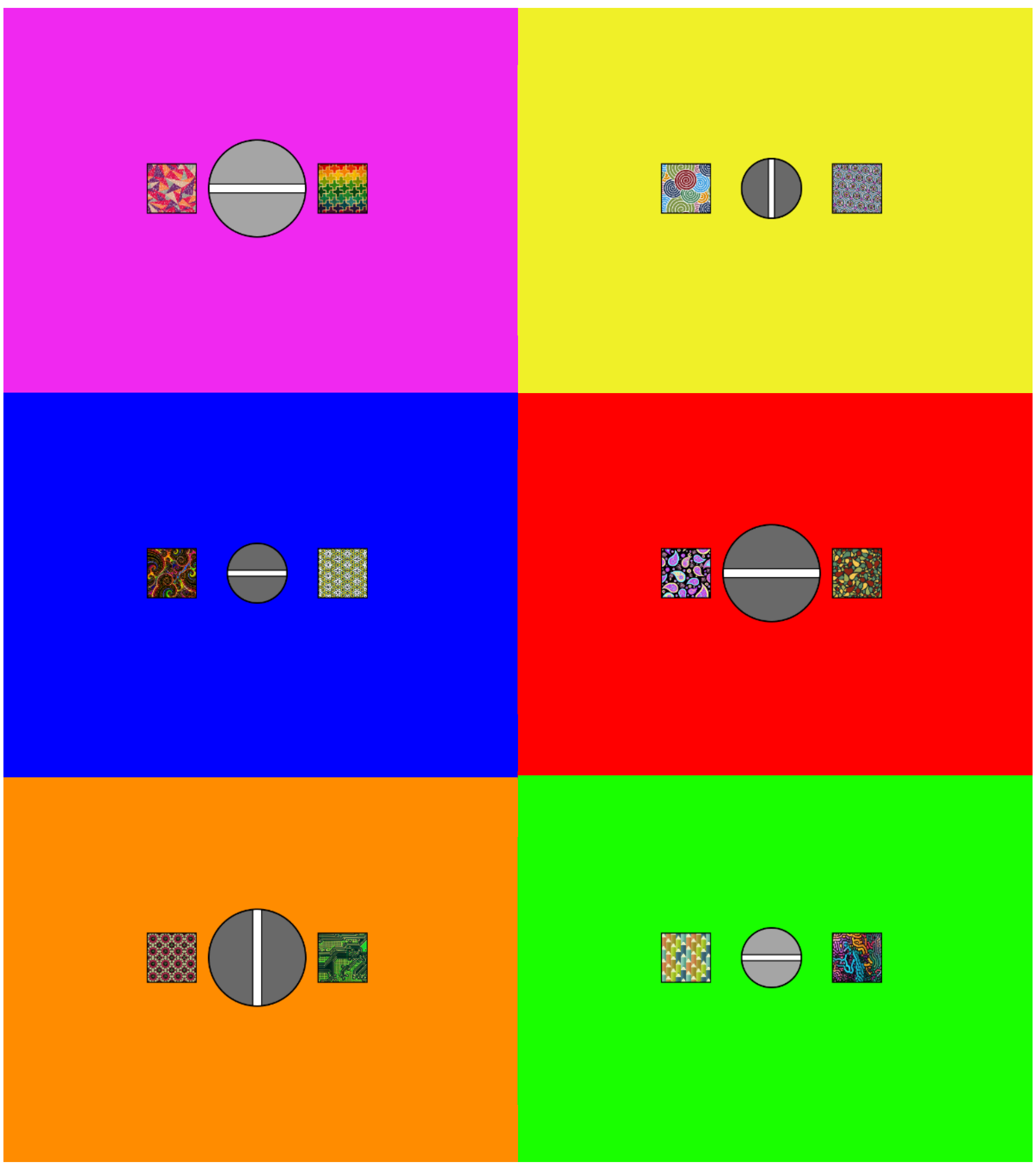
Victor M. Navarro, Ridhi, Jani, and Edward A. Wasserman

Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242.

E-mail: ed-wasserman@uiowa.edu

Examples of visual displays

The following are some displays exemplifying the possible combinations among training stimuli, colored backgrounds, and report buttons. Each of the eight training stimuli was equally likely to appear on all six backgrounds. Both background color and report buttons were unique to each of the six types of problems. As described in the text, the assignments differed from bird to bird.



Supplemental Material 2

Pigeon Category Learning: Revisiting the Shepard, Hovland, and Jenkins (1961) Tasks

Victor M. Navarro, Ridhi, Jani, and Edward A. Wasserman

Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242.

E-mail: ed-wasserman@uiowa.edu

Mixed-effects models information.

All mixed-effects models were fitted in R, using the lme4 package (https://cran.r-project.org/web/packages/lme4). Mixed-effect models are a type of hierarchical regression that allows the estimation of unbiased coefficients, by means of capturing the variance produced by nested factors (in this case, the variance produced by individual differences among our pigeons).

Although the fixed-effects structure of the model is determined by the experimental design itself, the random-effects structure that captures the nested variance needs to be identified through model selection. Model selection was performed in steps. Starting with a model that included pigeon intercepts, we compared models that had random-effects structures of increasing complexity. At each step, the feasibility of the more complex model was compared to its simpler counterpart by using χ^2 tests. If the χ^2 was significant, then the random-effect was kept in the final model's random-effect structure.

Logistic mixed-effects model for the Concurrent Phase data:

This model included the natural logarithm of session (1 to 180, centered) and structural ratio (centered) as fixed effects. Furthermore, the model included a random pigeon intercept and random pigeon slopes for the natural logarithm of session, structural ratio, and their interaction.

Logistic mixed-effects model for the Successive Phase data:

This model included the natural logarithm of session (1 to 30, centered) and structural ratio (centered) as fixed effects. Furthermore, the model included a random pigeon intercept and random pigeon slopes for the structural ratio and the Session × Structural Ratio interaction.

Supplemental Material 3

Pigeon Category Learning: Revisiting the Shepard, Hovland, and Jenkins (1961) Tasks

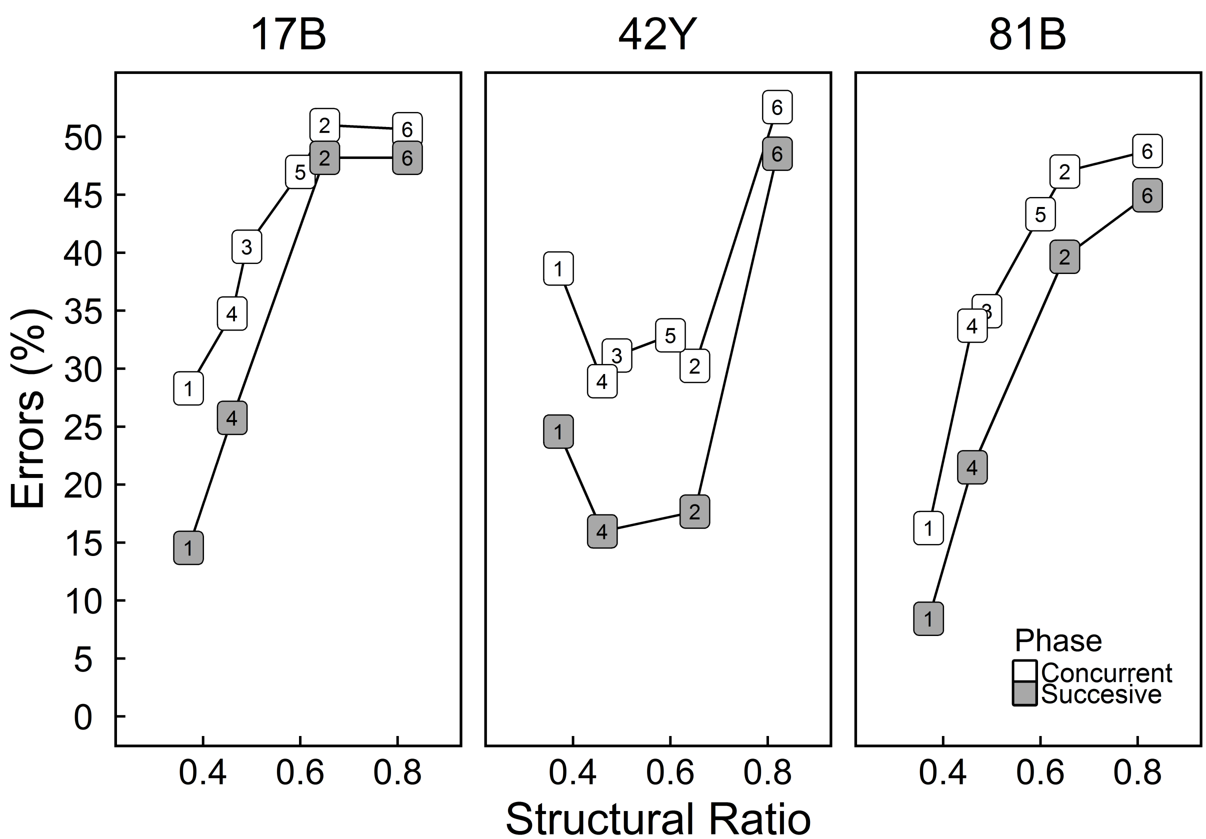
Victor M. Navarro, Ridhi, Jani, and Edward A. Wasserman

Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242.

E-mail: ed-wasserman@uiowa.edu

Individual pigeons' percentage of errors as a function of each task's structural ratio.

The figure below depicts each pigeon's overall percentage of errors in each task (1 to 6 in the concurrent phase, and 1, 2, 4, and 6 in the successive phase). As denoted in the figure, both 17B's and 81B's errors were linearly related to the structural ratio of each task. In contrast, 42Y's errors in the Type 2 task were almost as low as those in the Type 4 task.



Supplemental Material 4

Pigeon Category Learning: Revisiting the Shepard, Hovland, and Jenkins (1961) Tasks

Victor M. Navarro, Ridhi, Jani, and Edward A. Wasserman

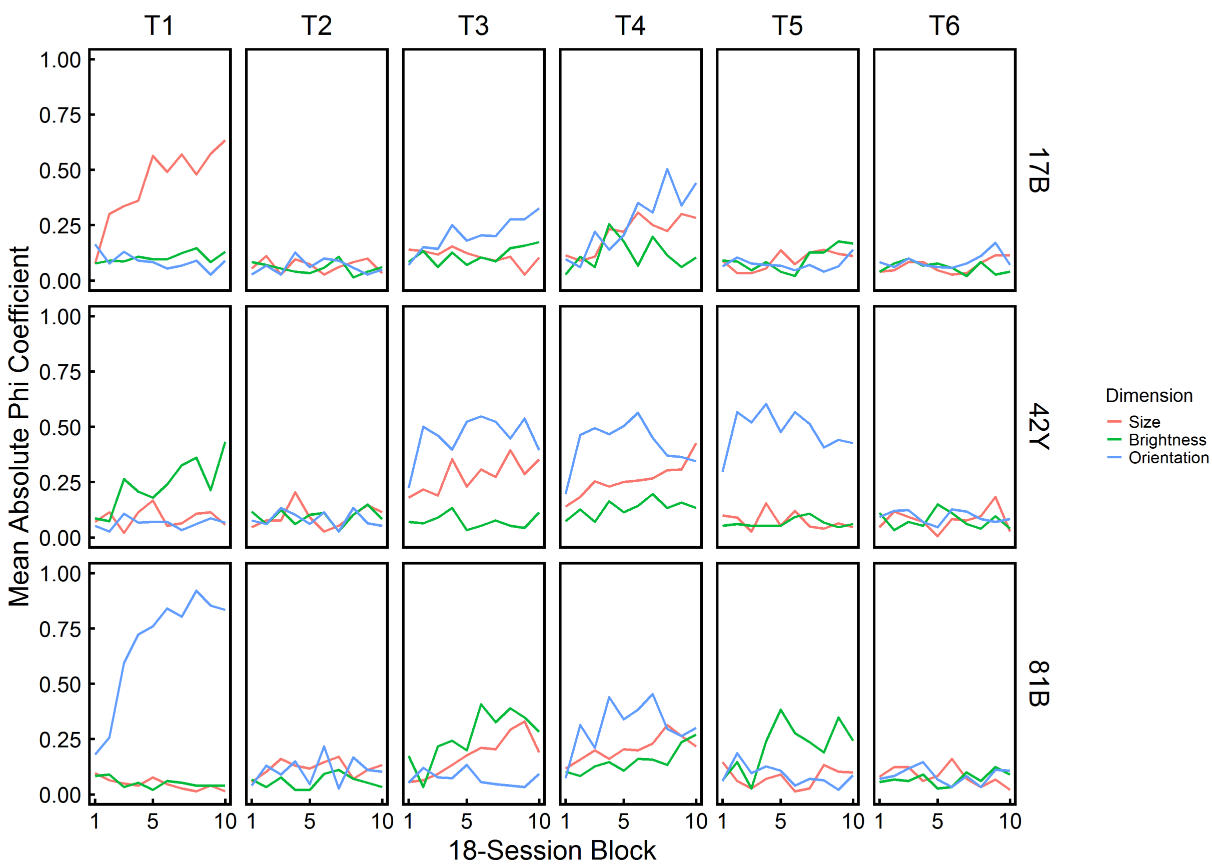
Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242.

E-mail: ed-wasserman@uiowa.edu

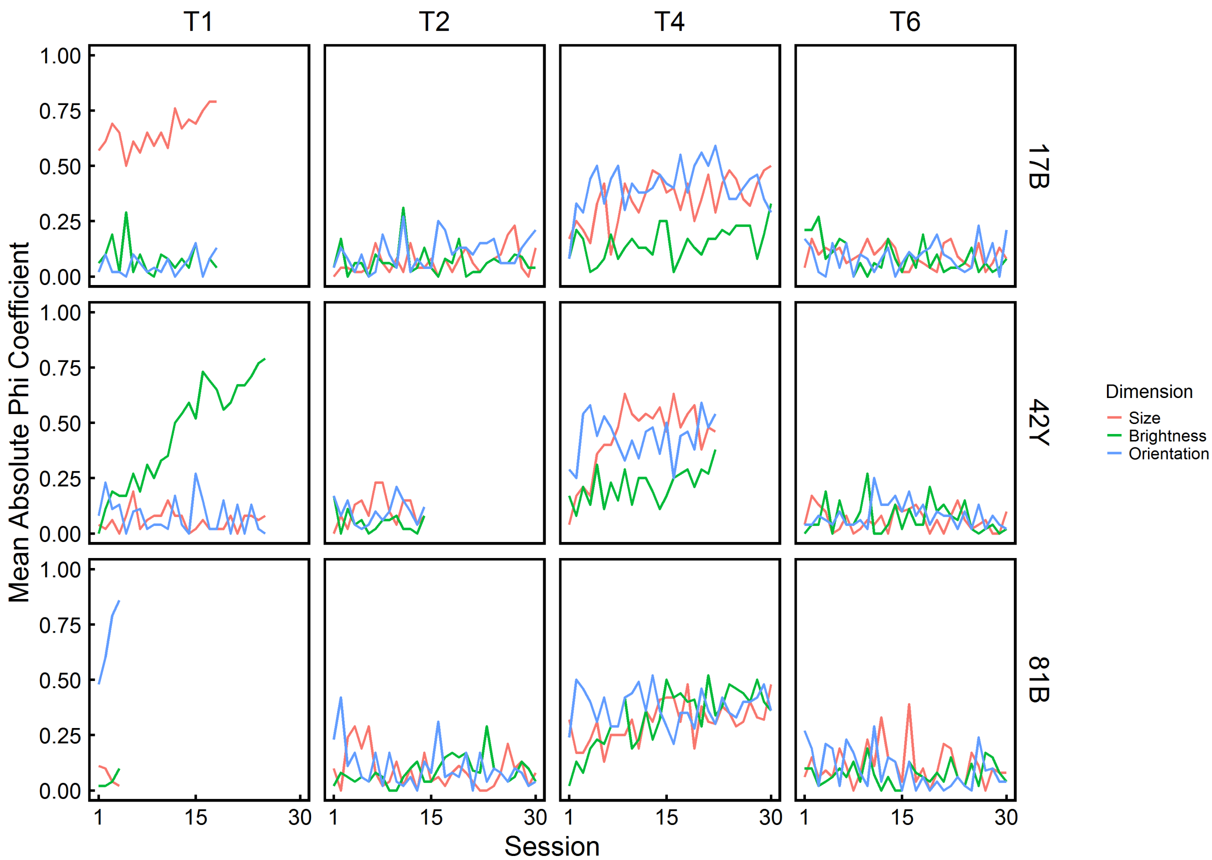
Phi coefficients between dimensional values and category choice.

The figures below depict each pigeon's mean absolute Phi coefficient between the values of each dimension and their category choices. In this setting, the absolute Phi coefficient indicates the extent to which the pigeons categorized the training stimuli on the basis of each dimension's values, with a Phi of 1 indicating a perfect relation between dimensional values and category choice (i.e., perfect unidimensional strategy) and 0 indicating no relation between them. Note that under perfect performance, only Type 1 will show a Phi of 1 along a single dimension. Due to their nonlinear structures, perfect performance in Types 2 and 6 will show a Phi of 0 along all three dimensions. Finally, perfect performance in Types 3, 4, and 5 will show a Phi of 0.5 along 2, 3, and 1 of the dimensions, respectively.

Concurrent phase:



Successive phase:



Supplemental Material 5

Pigeon Category Learning: Revisiting the Shepard, Hovland, and Jenkins (1961) Tasks

Victor M. Navarro, Ridhi, Jani, and Edward A. Wasserman

Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242.

E-mail: ed-wasserman@uiowa.edu

ALCOVE's fitting process

The comparison of pigeons' performance across the concurrent and successive phases suggested that concurrent programming did not materially alter our pigeons' performance on each task when trained separately. Therefore, we trained ALCOVE with one problem at a time. In doing so, the attentional weights for each dimension ($a_{brightness}$, $a_{orinetation}$, a_{size}) were all set to 1 before each task was trained (simulations that trained ALCOVE with all six tasks concurrently did not meaningfully alter the model's behavior). For all simulations, the dimensional values along each of the three dimensions comprising the stimuli were encoded as 0 or 1. Thus, ALCOVE's input layer contained three nodes. Except for "ALCOVE + dimensional discriminability" (see ahead), the exemplars in the hidden layer were activated as follows:

$$a_j^{hid} = \exp\left[-c\left(\sum_i \alpha_i \left|h_{ji} - a_i^{in}\right|^r\right)^{q/r}\right]$$

where a_j^{hid} is the activation of hidden node (or exemplar) j, α_i is the attentional weight for dimension i (brightness, orientation, or size, in the present case), h_{ji} is the dimensional value along dimension i for the hidden exemplar j, c is a global sensitivity parameter, and a_i^{in} is the dimensional value along dimension i for the presented stimulus. ALCOVE's hidden layer contained 8 nodes, after the number of unique stimuli presented in each task. For the current set of simulations, we fixed both q and r at 1 (i.e., city-block metric). ALCOVE's output category nodes are activated as follows:

$$a_k^{out} = \sum_{\substack{hid \\ j}} w_{kj} a_j^{hid}$$

where a_k^{out} is the activation of category node k, and w_{kj} is the response weight between hidden node j and category node k. Finally, the activation of category nodes is mapped onto response probabilities as follows:

$$P(K) = \exp\left(\phi a_K^{out}\right) / \sum_{\substack{out \\ k}} \exp\left(\phi a_k^{out}\right)$$

where P(K) is the probability of choosing category K, and ϕ is a parameter that scales the differences in the activation of the output nodes (i.e., controlling how deterministic responding is).

The "ALCOVE + dimensional learning" model expanded ALCOVE's single attentional learning rate parameter (λ_a) into three dimension-specific attentional learning rates ($\lambda_{brightness}$, $\lambda_{orientation}$, λ_{size}) in the equation determining attentional learning, as follows:

$$\Delta\alpha_i = -\lambda_i \sum_{\substack{hid \\ j}} \left[\sum_{\substack{out \\ k}} (t_k - a_k^{out}) w_k \right] a_j^{hid} c |h_{ji} - a_i^{in}|$$

where, λ_i is the attentional learning rate for dimension i, t_k is the teacher value for node k, a_k^{out} is the activation of output node k, and w_k is the response weight for node k.

The "ALCOVE + dimensional discriminability" model expanded ALCOVE's single sensitivity parameter into three dimension-specific sensitivity parameters ($c_{brightness}$, $c_{orientation}$, c_{size}) in the equation determining the activation of hidden nodes, as follows:

$$a_j^{hid} = \exp\left[-\left(\sum_i c_i \alpha_i |h_{ji} - a_i^{in}|^r \right)^{q/r} \right]$$

where c_i is the sensitivity parameter for dimension i. Additionally, the equation determining attentional learning was modified as follows:

$$\Delta\alpha_i = -\lambda_\alpha \sum_{\substack{hid \\ j}} \left[\sum_{\substack{out \\ k}} (t_k - a_k^{out}) w_k \right] a_j^{hid} c_i |h_{ji} - a_i^{in}|$$

where λ_α is a global attentional learning rate.

Each model was coded in MATLAB, and its parameters were optimized using a genetic algorithm (from MATLAB's global optimization toolbox). The parameters of each model were optimized to minimize the negative log likelihood of the model, given an individual pigeon's responses, on a trial-by-trial basis.

Supplemental Material 6

Pigeon Category Learning: Revisiting the Shepard, Hovland, and Jenkins (1961) Tasks

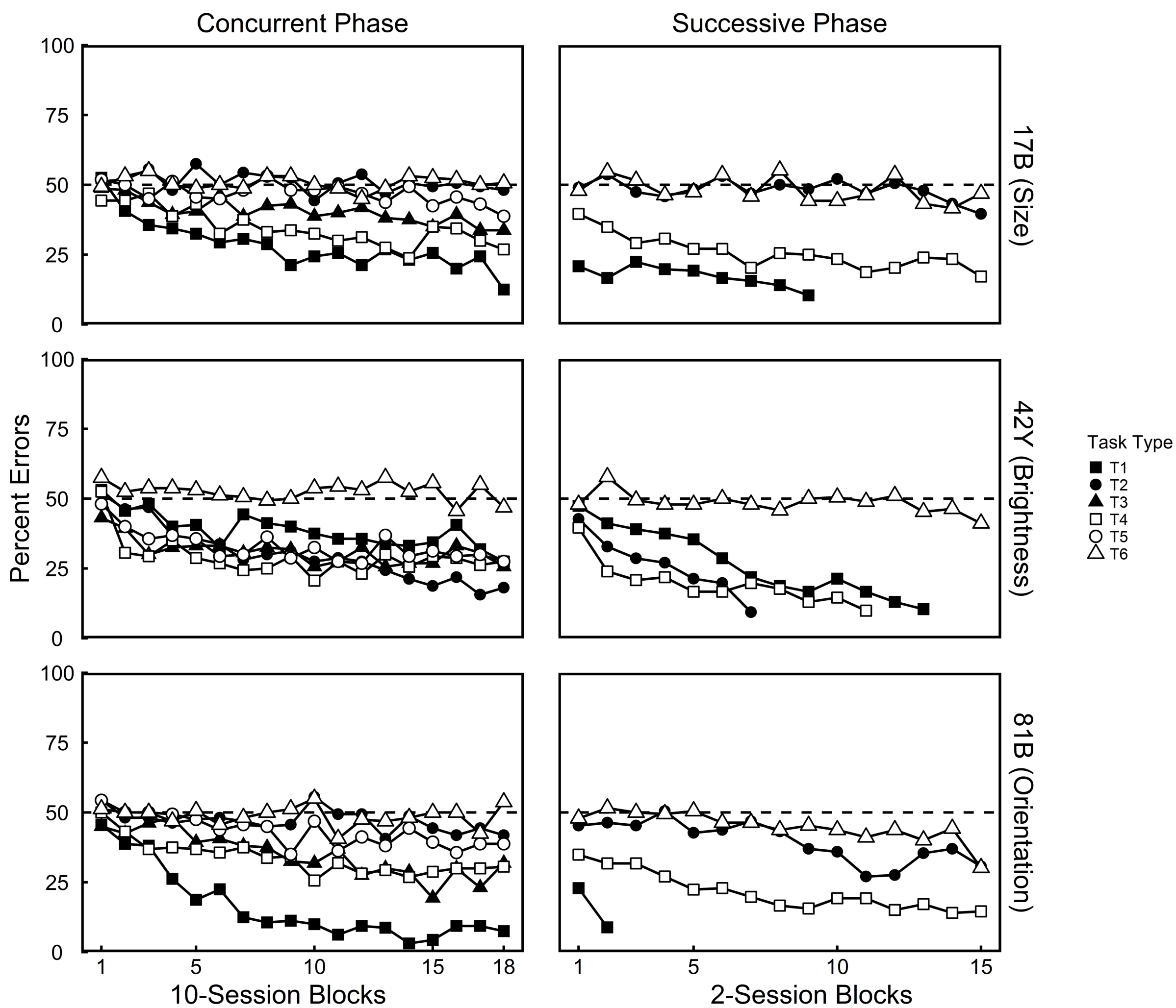
Victor M. Navarro, Ridhi, Jani, and Edward A. Wasserman

Department of Psychological and Brain Sciences, The University of Iowa, Iowa City, IA 52242.

E-mail: ed-wasserman@uiowa.edu

Learning curves for concurrent and successive phases

The following figure depicts the percentage of errors per task for each pigeon, across session blocks in the concurrent and successive phases (left and right columns, respectively). Note that no pigeon showed a sudden decrease in the percentage of errors on either the Type 1 task or the Type 2 task. Each bird name also denotes (in parentheses) the relevant dimension to determine category membership in the Type 1 task.



.