

Deferred Feedback Does Not Dissociate Implicit and Explicit Category-Learning Systems: Commentary on Smith et al. (2014)



Mike E. Le Pelley¹, Ben R. Newell¹, and Robert M. Nosofsky²

¹School of Psychology, UNSW Sydney, and ²Department of Psychological and Brain Sciences, Indiana University Bloomington

Received 6/27/18; Revision accepted 2/14/19

Proponents of the existence of distinct implicit and explicit category-learning systems have reported numerous dissociations consistent with this dual-systems account (for a review of this extensive literature, see Ashby & Maddox, 2011). However, research has failed to replicate these findings, or confounds admit alternative explanations (e.g., Dunn, Newell, & Kalish, 2012; Edmunds, Milton, & Wills, 2015; Kalish, Newell, & Dunn, 2017; Newell, Dunn, & Kalish, 2010, 2011; Nosofsky & Kruschke, 2002; Stanton & Nosofsky, 2013).

Here, we examine a study by Smith et al. (2014), who claimed to have found “one of the strongest explicit-implicit dissociations yet seen in the categorization literature” (p. 447). Their participants learned to categorize stimuli varying on two dimensions. For some participants, optimal performance required learning about only one dimension; others had to integrate information about both dimensions. As an example, consider line stimuli varying in length and angle (Fig. 1a). In a unidimensional, vertical structure, short lines belong to Category A and long lines to Category B; angle has no bearing on classification. For a diagonal structure, both dimensions influence the correct response: Optimal performance requires a decision based on perceptual information integrated across dimensions.

Smith et al. contend that the vertical structure is learned via an explicit rule-based process subserved by a declarative system, whereas the diagonal structure is the domain of an implicit, procedural system that performs information integration. According to this account, diagonal structures lie beyond the explicit, rule-based system because the optimal strategy is difficult to describe verbally (Ashby & Valentin, 2017). This idea motivated Smith et al.’s critical manipulation of feedback (which they also referred to as a manipulation of reinforcement). On each training trial, participants in

an immediate-feedback condition saw the stimulus, made a categorization response, and were told whether their response was correct. Participants in a deferred-feedback condition instead responded to blocks of six stimuli in a row before receiving feedback stating how many of those six responses were correct; but they were not told which stimuli they had responded to correctly and which incorrectly.

As described by Smith et al. (p. 451), a crucial result was an interaction between category structure (vertical vs. diagonal) and feedback (immediate vs. deferred) on categorization accuracy. Whereas deferring feedback had little effect (relative to immediate feedback) on accuracy for a vertical structure, it significantly impaired diagonal-structure performance. To explain this interaction, Smith et al. argued that deferring feedback selectively disables the implicit category-learning system because when feedback is not timely, the temporally constrained associative-learning process that embodies this system cannot operate. According to this account, vertical-structure learning is unaffected by deferring feedback because it relies on a verbal rule, and success can be evaluated “at block’s end just as at trial’s end” (p. 450). In contrast, diagonal learning is disrupted because deferring feedback disables the implicit-learning process required for optimal performance, forcing participants to fall back on the explicit system that treats each dimension separately.

We offer an alternative interpretation based on differences in cognitive complexity and memory demands

Psychological Science
1–7

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0956797619841264

www.psychologicalscience.org/PS



Corresponding Author:

Mike E. Le Pelley, UNSW Sydney, School of Psychology, Sydney, New South Wales 2052, Australia

E-mail: m.lepelley@unsw.edu.au

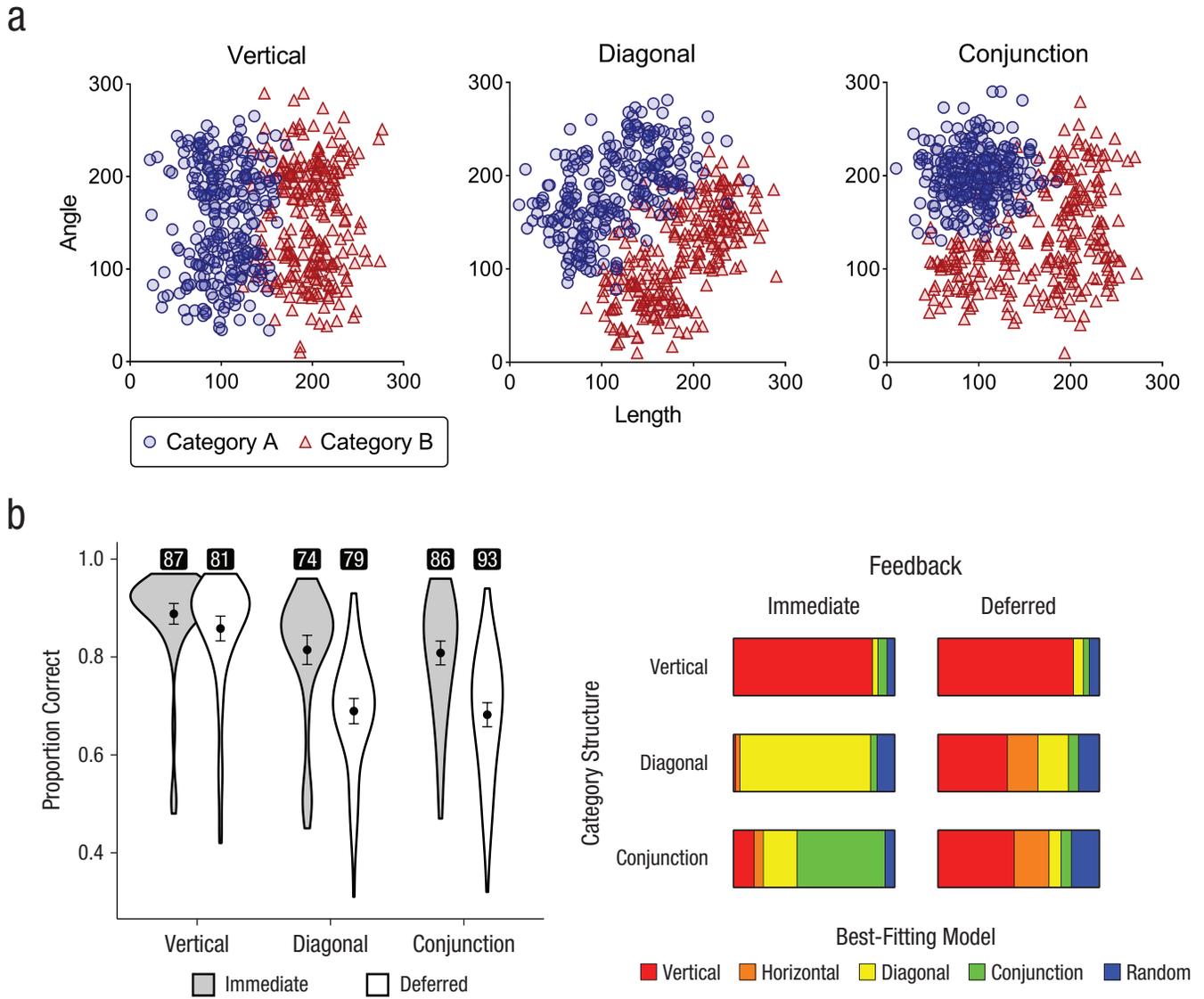


Fig. 1. Category structures and experimental results. Examples of the vertical, diagonal, and conjunction category structures used in the current experiment are shown in (a). Values on each dimension relate to the length and angle of a line that participants were required to classify as belonging to Category A or Category B. Values are in arbitrary units; see Table S1 in the Supplemental Material available online for further information. Results for responses over the last 100 categorization trials are shown in (b). Violin plots (left) show the proportion of correct categorization responses for participants trained with vertical, diagonal, and conjunction category structures and who were given either immediate or deferred feedback. The shape of each plot shows the probability density of the data at different levels of proportion correct; the central point and error bars show the mean and 95% confidence interval, respectively. Values above each plot show number of participants per condition. The stacked bars on the right show the proportion of participants for whom each type of decision-bound model (vertical, horizontal, diagonal, conjunction, or random guessing) provided the best fit to categorization responses in the last 100 trials. Results are shown separately for each combination of category structure and type of feedback that participants were trained with.

(Nosofsky, Stanton, & Zaki, 2005). Optimal performance with a vertical structure entails a decision strategy referring to a single dimension value. Optimal diagonal-structure performance instead requires that, for each stimulus, the observer combine two separable dimension values and remember each unique combination. Given the relative ease and low memory demands associated with the vertical structure, it is unsurprising that deferred feedback has little impact on performance. By

contrast, it seems reasonable that deferring feedback will have a major negative impact on performance in the demanding diagonal-categorization task.

To decide between these interpretations—multiple systems versus cognitive demands—we introduced a *conjunction* category structure to the design (Fig. 1a). Like the diagonal categorization, the conjunction categorization requires learning about two dimensions. Critically, however, unlike the diagonal structure, the

conjunction structure is viewed by multiple-system theorists as a rule-based task that relies on the explicit system. This is because the observer makes separate decisions about values on each dimension and then combines those decisions to make a categorization (“short length and large angle → Category A; otherwise Category B”) rather than integrating perceptual information across dimensions as for the diagonal task. We emphasize that in numerous previous studies, multiple-systems theorists themselves have used conjunction structures as major examples of what they theorize are rule-based tasks (e.g., Filoteo, Lauritzen, & Maddox, 2010; Helie & Ashby, 2012; Maddox, Bohil, & Ing, 2004).

According to the multiple-systems account articulated by Smith et al., conjunction-task performance should be unaffected by deferring feedback because conjunction structures are learned by an explicit rule-based system, and deferring feedback does not affect rule-based learning. Instead, as theorized by Smith et al., rule-based learning “would flourish under deferred reinforcement” (p. 450). By contrast, under the cognitive-demands interpretation, conjunction performance should suffer from deferred feedback just as for the diagonal task: Clearly, the cognitive complexity and memory demands of the conjunction structure are greater than for the vertical structure.

Method

Participants

This experiment was run online via Amazon Mechanical Turk. A demonstration version of the task can be accessed at <http://unsw-mlp-deffb.appspot.com?cat=1&fb=1>. A total of 500 participants (235 females; age: $M = 36.9$ years, $SEM = 0.5$) were randomly assigned to category structure (vertical, conjunction, diagonal) and feedback (immediate, deferred) conditions; Figure 1b shows the sample size in each condition. Our sample size of approximately 80 per condition was significantly larger than that of Smith et al. (21 per condition), and it gave us a power of .98 to detect an interaction between category structure and feedback type with an effect size (η_p^2) of .048 (the effect size observed by Smith et al.). Each participant received \$5 for completing the task (which took ~30 min). The best-performing half of participants also received a performance-related bonus of \$3 (see below). This study was approved by the Human Research Ethics Advisory Panel (Psychology) of UNSW Sydney.

Apparatus, stimuli, and procedure

The multiple-systems theory is intended to apply generally to the class of separable-dimension stimuli. The

stimuli we used here were lines varying in length and angle. These are classic examples of separable-dimension stimuli and have been used in numerous previous experiments published by multiple-systems theorists (e.g., Ashby, Ell, & Waldron, 2003; Filoteo et al., 2010). Smith et al. used alternative stimuli defined by the area of a rectangle and the density of green pixels within the rectangle (but did not provide a reason for this choice). Although this possibility is not the central theme of this Commentary, we were concerned that such stimuli may give rise to a salient emergent dimension based (for example) on the total number of green pixels in the display. If so, the psychological category structures learned by some participants might not correspond to those intended by the experimenters. We decided to use the line stimuli to avoid this potential difficulty, since there is less potential for salient emergent dimensions to arise for these stimuli.

Stimulus presentation was controlled by jsPsych software (de Leeuw, 2015). Each stimulus was a single blue line presented on a white background. Table S1 in the Supplemental Material available online details the procedure used to generate the stimuli, and Figure 1a shows examples of the resulting category structures.

Participants were informed that on each trial they would see a line and that their task was to decide whether it belonged to Category C or Category M by pressing either the C or M key, as appropriate. For some participants—chosen randomly—Category A (for stimulus generation) was mapped onto Category C (for responses) and Category B was mapped onto Category M; for the remaining participants, this was reversed. Participants were told that they should try to make as many correct responses as possible and that once the experiment was complete, they would receive a \$3 bonus if their accuracy was above average compared with that of the other participants. Participants in the immediate-feedback condition were informed that they would be told after each response whether it was correct. Participants in the deferred-feedback condition were informed that after they had responded to six lines, they would be told how many of those six responses were correct, “but you won’t be told which individual responses were correct and which were incorrect.” Check questions were used to verify that participants had understood all instructions before trials began.

Participants completed 17 sets of 30 trials each (510 trials total). On each trial, the line stimulus was presented centrally inside a white square with a black border (side length = 300 pixels) until the participant made a response. Participants in the immediate-feedback condition saw a message saying either “correct” or “incorrect” (presented centrally for 800 ms), followed by a blank interval of 800 ms before the next

trial began. In the deferred-feedback condition, for the first five trials in each six-trial block, each response was followed by the next stimulus after a blank interval of 833 ms. The sixth response was followed by the feedback “You scored X out of 6” (displayed for 4,800 ms), where X was the number of correct responses in the previous block. The next six-trial block then began after a blank interval of 1,000 ms. All participants took a short break after each set of 30 trials.

Data analysis and formal modeling

Following Smith et al., we analyzed participants’ categorization accuracy over the final 100 trials as a function of category structure (vertical, diagonal, conjunction) and feedback (immediate, delayed). We also modeled participants’ category choices over the final 100 trials to investigate the classification strategy that they had adopted and, in particular, how this strategy was influenced by deferring feedback. We fitted five different models to each participant’s data. Four of the models differed according to where hypothesized category boundaries lay in the stimulus space shown in Figure 1a: (a) vertical boundary; (b) horizontal boundary; (c) diagonal boundary; and (d) conjunction, given by combining a vertical and a horizontal boundary. The fifth model accounted for random guessing. We used a maximum-likelihood criterion to estimate the parameters of each model (e.g., what is the x -value of the vertical boundary through stimulus space that best partitions a participant’s “Category A” and “Category B” responses?). We then used the Bayesian information criterion (BIC; Schwarz, 1978) to determine the best-fitting model for each participant. The BIC penalizes more complex models by including a penalty term based on the number of parameters in the model.

Results

Accuracy-based analyses

Experiment code and raw data from this experiment are available at <https://osf.io/6s8tc>. Figure 1b shows mean accuracy over the final 100 trials in each condition. A 3 (category structure: vertical, diagonal, conjunction) \times 2 (feedback: immediate, delayed) analysis of variance (ANOVA) revealed main effects of category structure, $F(2, 494) = 65.7, p < .001, \eta_p^2 = .21$, and feedback, $F(1, 494) = 82.5, p < .001, \eta_p^2 = .14$, as well as a significant interaction, $F(2, 494) = 9.62, p < .001, \eta_p^2 = .037$. Šidák-corrected pairwise t tests (critical $\alpha = .017$) revealed that deferring feedback significantly impaired accuracy for the diagonal structure, $t(151) = 6.33, p < .001, d = 1.02$, and for the conjunction structure, $t(187)$

$= 7.25, p < .001, d = 1.08$, but not for the vertical structure, $t(166) = 1.82, p = .071, d = 0.28$.

To decompose the interaction revealed by the omnibus ANOVA, we ran follow-up 2×2 ANOVAs to compare the effect of feedback between each pair of category structures. For the 2×2 analysis of vertical and diagonal conditions, we found a significant Category Structure \times Feedback interaction, $F(1, 317) = 13.8, p < .001, \eta_p^2 = .042$, indicating that deferred feedback caused a greater impairment in the diagonal condition than the vertical condition. This replicates the previous finding of Smith et al. (despite some differences in stimuli and procedure) and was a key dissociation that they pointed to as evidence for dissociable category-learning processes. To recap, they argued that learning a diagonal structure requires an implicit, associative-learning system in order to integrate information across the two stimulus dimensions, and that this associative-learning system cannot operate when feedback is deferred.

Critically, however, the 2×2 analysis of the vertical versus conjunction conditions also yielded a significant interaction, $F(1, 353) = 16.0, p < .001, \eta_p^2 = .045$, with greater feedback-related impairment in the conjunction condition than the vertical condition. That is, we observed the same feedback-related dissociation for the conjunction condition as we did for the diagonal condition. Analysis of the diagonal and conjunction conditions did not yield a significant interaction, $F(1, 338) = 0.002, p = .96, \eta_p^2 < .001$. In other words, the magnitude of the effect of deferring feedback did not differ significantly for participants trained with a diagonal versus a conjunction structure. Evidence for this null effect is bolstered by a Bayesian statistical analysis reported in our Supplemental Material available online, with a Bayes factor of 45.9 in favor of the null interaction. This lack of a difference in the magnitude of the deferred-feedback effect across the diagonal and conjunction conditions is not central to our lines of argument, because we make no claim that those two conditions are exactly equated in terms of their overall cognitive complexity and memory demands. Nevertheless, the result severely challenges the multiple-systems account of the deferred-feedback effects, because (according to this account) the diagonal condition is an information-integration task and the conjunction condition a rule-based task. That is, this experiment shows that information-integration structures are not necessary to produce an impairment by deferred feedback.

One further detail of our findings is notable: Performance was significantly better for the vertical task than the diagonal task when immediate feedback was given, $t(159) = 4.10, p < .001, d = 0.65$. This finding is consistent with decades of research showing that the diagonal

structure is harder for humans to learn (e.g., Ashby et al., 2003; Kruschke, 1993; Newell et al., 2010; Nosofsky et al., 2005). However, in Smith et al.'s study, vertical- and diagonal-task performance did not differ significantly when immediate feedback was given in the final 100 trials of training, and on this basis, they argued that task difficulty was matched in their study. We note, though, that mean accuracy for their vertical task was 5% higher than for their diagonal task. This (nonsignificant) difference was similar to the corresponding vertical-versus-diagonal difference in our experiment (7.3%), which was significant, most likely as a consequence of our study's much greater statistical power. Smith et al.'s study (with only 21 participants per condition) had a power of only .54 to detect a vertical-versus-diagonal difference of this size ($d = 0.65$), so their null result under immediate feedback does not license the conclusion that task difficulty was matched.¹

Model-based analyses

Plots of individual participants' categorization responses labeled with the best-fitting model are available at <https://osf.io/6s8tc>. Visual inspection of these plots suggests that our model-fitting procedure was valid (e.g., for participants for whom the diagonal model was found to provide the best fit, the plot of Category A/B responses typically shows a clear diagonal division).

Figure 1b shows the proportion of participants in each condition for whom each model provided the best fit. For participants trained with immediate feedback, the modal best-fitting model was the appropriate model for that category structure: For participants trained with a vertical structure, the vertical model performed best overall; for the diagonal task, the diagonal model performed best; and for the conjunction task, the conjunction model performed best. By contrast, under training with deferred feedback, a unidimensional model performed best regardless of category structure. The finding that, for the diagonal condition, deferring feedback promoted use of a unidimensional strategy replicates the findings of Smith et al. But, crucially, our data show this pattern is not specific to information-integration category structures, as deferring feedback had the same effect for our conjunction condition (which, according to the multiple-systems account, is rule based). That is, participants faced with a difficult, cognitively demanding diagonal or conjunction structure tended to fall back on a unidimensional (vertical/horizontal) strategy when feedback was deferred. Consistent with this characterization, results revealed that the distribution of best-fitting models for the diagonal and conjunction conditions was significantly different when feedback was immediate, $\chi^2(4) = 71.1, p < .001$, but did not differ

significantly when feedback was deferred, $\chi^2(4) = 5.27, p = .26$. Contrary to Smith et al.'s claim that deferred feedback "sharply dissociates" information-integration and rule-based categorization systems, our results show that the effect of deferring feedback on participants' pattern of classification is not selective to tasks that (according to the multiple-systems account) require information integration, because exactly the same pattern was seen for a structure which (according to this account) was rule based. In both cases, when feedback was deferred, participants fell back on a one-dimensional strategy. The finding that many participants use a one-dimensional strategy as a basis for classification when they fail to solve the intended categorization problem does not, in our view, imply that some separate system is involved. Instead, it seems natural that participants would resort to a simple strategy entailing low cognitive demands when they fail to solve a highly demanding task.

Discussion

Smith et al. argued that deferring feedback specifically impairs diagonal-structure performance because it disables the implicit associative system required for information-integration learning but leaves an explicit rule-based system unhindered. Contrary to this suggestion, our results showed that impairment when feedback is deferred is not specific to diagonal, information-integration structures but also occurs for conjunction structures that (according to the multiple-systems view) are rule based. These findings do not follow from Smith et al.'s multiple-systems account but follow naturally from a cognitive-demands account: The cognitive complexity and memory demands of diagonal and conjunction tasks are greater than for the vertical task, so deferring feedback will impair both two-dimensional tasks and may drive participants to a less-demanding unidimensional strategy.

A multiple-systems theorist might try to dismiss our findings because we violated methodological procedures for investigating "dissociations." For example, Smith et al. argued that their vertical and diagonal tasks were equated in difficulty for participants who received immediate feedback, and it was only deferred feedback that impaired diagonal-task performance. However—as noted above—their failure to find a significant difference in difficulty when feedback was immediate was likely a consequence of their study being underpowered. Smith et al. advanced auxiliary lines of argument to support their claim of equal task difficulty when feedback was immediate. We address these arguments in detail in the Supplemental Material. The most important point is that the general claim that the vertical and

diagonal structures are matched in difficulty conflicts with decades of empirical research, and our own data yield the ubiquitous result that the diagonal task is more difficult. Notwithstanding issues regarding statistical power, a more limited claim might be made that task difficulty was matched in the special case in which Smith et al.'s rectangle-pixel stimuli were used; however, Smith et al. provide no explanation of what aspect of their stimuli might yield that highly unusual result.

One approach to replying to our Commentary is to hypothesize that effects of deferred feedback will depend on both the implicit or explicit nature of the task as well as the cognitive complexity of the task to be solved. This is a reasonable possibility that should be pursued in future research. However, it is not the hypothesis advanced by Smith et al., who made no allowance for a role of cognitive complexity in mediating the effects of deferred feedback; indeed, they instead advanced the untenable general claim that the vertical and diagonal tasks were matched in difficulty.

Importantly, we are not denying the interest value or even the plausibility of the hypothesis that separate cognitive neural systems—governed by different operating characteristics—mediate learning of different categorization tasks. Instead, our concern is with the evidence base used to bolster the hypothesis. Smith et al.'s manipulation of rule-based versus information-integration structure was confounded with factors relating to cognitive complexity and task difficulty, and the current experiment provides clear evidence that these factors influence the effects of interest. Thus, our study adds to the growing body of work questioning the dissociation logic that has been used thus far to bolster the multiple-category-systems hypothesis.

Finally, a limitation of our Commentary is that we have not provided rigorous definitions of the interrelated constructs of cognitive complexity, memory demands, and task difficulty. Our goal, however, is not to develop a complete theory of the many factors that may modulate the effect of deferred feedback on categorization performance. Instead, our concern is that Smith et al. attributed *all* of the effects to whether the task was rule based (explicit) or information-integration based (implicit) and claimed that they had produced one of the strongest explicit-implicit dissociations yet seen in the categorization literature. We believe that their conclusions were overstated.

Action Editor

Marc J. Buehner served as action editor for this article.

Author Contributions

All three authors contributed equally to this work. All authors developed the study concept and contributed to the study

design. Data were collected and analyzed by M. E. Le Pelley and R. M. Nosofsky. All authors interpreted the findings, drafted and revised the manuscript, and approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

R. M. Nosofsky was supported by National Science Foundation Grant 1534014.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797619841264>

Open Practices



Deidentified data from this experiment and all materials for the experiment (including code used for procedural generation of the stimuli) have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/6s8tc>. A demonstration version of the task can be accessed at <http://unsw-mlp-deffb.appspot.com?cat=1&fb=1>. The procedures used for generating all materials are fully described in the article. The new experiment reported in this article was not preregistered. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797619841264>. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

Note

1. Because Smith et al.'s raw data were not available at the time this Commentary was written, we could not conduct Bayesian tests of strength of support for their null hypothesis that the vertical and diagonal tasks were equally difficult. Nor could we compare performance across the two tasks earlier in training, prior to the final 100 trials.

References

- Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, *31*, 1114–1125.
- Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.
- Ashby, F. G., & Valentin, V. V. (2017). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Handbook of*

- categorization in cognitive science* (pp. 157–188). San Diego, CA: Academic Press.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.
- 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.
- Edmunds, C. E. R., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *Quarterly Journal of Experimental Psychology*, *68*, 1203–1222.
- 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.
- Helie, S., & Ashby, F. G. (2012). Learning and transfer of category knowledge in an indirect categorization task. *Psychological Research*, *76*, 292–303.
- Kalish, M. L., Newell, B. R., & Dunn, J. C. (2017). More is generally better: Higher working memory capacity does not impair perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 503–514.
- Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, *5*, 3–36.
- Maddox, W. T., Bohil, C. J., & Ing, A. D. (2004). Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review*, *11*, 945–952.
- Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory & Cognition*, *38*, 563–581.
- Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 167–215). San Diego, CA: Academic Press.
- Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*, *9*, 169–174.
- Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition*, *33*, 1256–1271.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
- Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred feedback sharply dissociates implicit and explicit category learning. *Psychological Science*, *25*, 447–457.
- Stanton, R. D., & Nosofsky, R. M. (2013). Category number impacts rule-based and information-integration category learning: A reassessment of evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1174–1191.