

# SUSTAIN captures category learning, recognition, and hippocampal activation in a unidimensional vs information-integration task

Anonymous CogSci submission

## Abstract

There is a growing interest in alternative explanations to the dual-system account of how people learn category structures varying in their optimal decision bounds (unidimensional and information-integration structures). Recognition memory performance and hippocampal activation patterns in these tasks are two interesting findings, which have not been formally explained. Here, we carry out a formal simulation with SUSTAIN (Love, Medin, & Gureckis, 2004), an adaptive model of category learning, which had great success in accounting for recognition memory performance and fMRI activity patterns. We show, for the first time, that a formal single-system model of category learning can accommodate recognition performance after learning and is consistent with fMRI data obtained while learning these structures.

**Keywords:** categorization; recognition memory; formal model; SUSTAIN

## Introduction

One commonly used pair of category structures in categorization research includes the unidimensional (UD) and information-integration (II) category structures. The shared characteristic of UD and II is that they are perceptually the same, but differ in their optimal decision bounds. These properties were initially used for separating perceptual processes encoding the visual information from decision processes assigning a category response to the perceptual effects (Ashby & Gott, 1988). Figure 1 shows how stimuli varying in size and brightness are distributed within these two category structures on either side of the boundaries. UD category structures have a vertical decision bound: if the square is darker than  $x$ , then it is category A, otherwise it is category B. Figure 1 shows that this optimal decision bound is parallel to one of the dimensional axis in the psychological space. II structures are defined by diagonal optimal decision bounds. Figure 1 shows that II decision bounds follow a linear function, where the gradient is neither zero, nor infinite.

Many experiments utilised these structures (e.g. Ashby & Valentin, 2018; Carpenter, Wills, Benattayallah, & Milton, 2016; Casale, Roeder, & Ashby, 2012; Donkin, Newell, Kalish, Dunn, & Nosofsky, 2015; Dunn, Newell, & Kalish, 2012; Nomura et al., 2007; Nosofsky, Stanton, & Zaki, 2005; Kalish, Newell, & Dunn, 2017; Le Pelley, Newell, & Nosofsky, 2019). Many initial empirical results were taken as evidence for COVIS (Ashby, Paul, & Maddox, 2011) — one formalization of the dual-system theory of categorization. Traditionally, dual-system theories have distinct architectures us-

ing functionally different mechanisms. The explicit system uses rules that can be easily verbalized, while the implicit system maps perceptual input onto category responses. COVIS considers UD and II to result from different decision processes part of the two distinct architectures, which selectively engage with the perceptual input. In COVIS, the explicit system engages with UD structures and the implicit procedural system engages with II structures.

However, results from multiple labs pointed out flaws in the experimental designs (Newell, Moore, Wills, & Milton, 2013) with potential alternative explanations (Le Pelley et al., 2019; Donkin et al., 2015; Dunn et al., 2012) or problems with the decision-bound analysis themselves (Edmunds, Milton, & Wills, 2018; Edmunds, Wills, & Milton, 2019; Edmunds, Inkster, Jones, Milton, & Wills, 2020). The way COVIS explains how people learn II structures were also questioned by Carpenter et al. (2016) and Edmunds, Wills, and Milton (2016), who provided direct evidence for an involvement of similar processes in both II and UD problems.

Carpenter et al. (2016) found that the medial temporal lobe (MTL) and specifically the hippocampus (HPC) were more active in the category learning task involving II structures compared to the task with UD structures. This MTL and HPC activity suggests either the involvement of the same system or a shared process. As MTL and HPC is thought to be essential for explicit memory, this finding suggest that people should have explicit access to category knowledge in II to a greater extent than in UD structures. This prediction contradicts the COVIS account on how II structures are learned. Edmunds et al. (2016) directly tested whether people have conscious access to information in II structures by including a recognition task in both UD and II problems. Including a recognition task directly measures the extent which people have access to category knowledge in either structures - and to what extent the neuroimaging data aligns with behavioural results. Edmunds et al. (2016) found better recognition memory after learning II compared to UD structures, essentially supplementing the neural data.

Building on these findings, we further supplement behavioural and neural data with evidence from computational modelling. Here, we provide a formal single-system explanation of these results. We do so by using SUSTAIN (Love et al., 2004), which in bundle with the neural and behavioural data will encompass all levels of analysis of II and UD prob-

lems (Marr & Poggio, 1979).

SUSTAIN is a model in a single-system framework able to accommodate a wide range of behavioural and neural phenomena (e.g. Love & Medin, 1998b, 1998a; Love, Markman, & Yamauchi, 2000; Love et al., 2004; Gureckis & Love, 2002, 2003, 2004; Davis, Love, & Preston, 2012; Mok & Love, 2019). This breadth is particularly admirable, because modelers tend to focus on a small subset of effects (Wills, O’Connell, Edmunds, & Inkster, 2017).

There are two reasons for using SUSTAIN. First, SUSTAIN could accommodate recognition memory performance in multiple tasks (Love & Gureckis, 2007; Davis et al., 2012; Mack, Love, & Preston, 2018). Second, SUSTAIN’s concept-forming and -altering mechanism, adaptive clustering, maps to HPC and MTL functions and activations (Love & Gureckis, 2007).

SUSTAIN views the hippocampus as the constructor and editor of clusters — binding information together into category representations, and views the MTL familiarity signals as indicators of cluster re-activations. These views have been reinforced by connecting computational modelling to neural activity patterns. For example, during rule-plus-exception learning, SUSTAIN makes trial-by-trial prediction about item recognition and error correction. These predictions directly and consistently mapped to MTL activations (Davis et al., 2012). Furthermore, SUSTAIN’s cluster-updating mechanism parallels HPC activity in response to changing task demands. SUSTAIN accommodated behavioural responses and HPC activity in subsequent learning tasks where the stimuli remained perceptually the same (Mack, Love, & Preston, 2016), while irrelevant features in the first task became essential in the new categorization problem. SUSTAIN’s behaviour and HPC activity both correlated how stored category representations had to be reorganized as a function of changing conceptual knowledge. SUSTAIN is well matched with how the HPC binds together information into meaningful category representations and updates the stored representations to match with goal-oriented changes in task-demands (for a complete review, see Mack et al., 2018). This logic leads to two predictions for SUSTAIN: (1) increased HPC activity in II should correspond with a higher number of clusters being recruited in II, because the more difficult task requires the binding and storing of larger sets of information into clusters than the simpler task (Love & Gureckis, 2007); (2) higher MTL and HPC activity and better recognition memory in II than in UD should correspond with SUSTAIN recruiting higher number of clusters and recognizing more items. In this paper, we test these prediction by simulating an experiment similar to Edmunds et al. (2016) with SUSTAIN.

## SUSTAIN

We refer the reader to Love et al. (2004) for the full description of the model’s architecture and Love and Gureckis (2007) for the full description of the supplementing architecture capturing recognition memory.

Briefly, SUSTAIN is an adaptive clustering model,

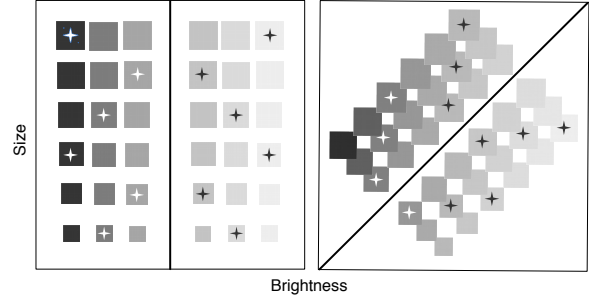


Figure 1: Representations of the category structures used in Edmunds et al. (2016). Left: Unidimensional, Right: Information-integration. The stars (colour irrelevant, not shown to the participants) mark possible stimuli that were removed during the training phase of the experiment.

which proposes that clusters underlie category representations (Love, 2005). SUSTAIN starts with one cluster, centred on the first input representation it encounters by default. When SUSTAIN encounters a stimulus, it computes similarity from all stored cluster representations in the psychological space. First, the distance is calculated for each dimension, then differentially weighted in the cluster activation function by attentional tunings. So similarity on dimensions with higher attentional tunings will be more impactful on which cluster is activated. The winning cluster will be the one with the highest activation. After this algorithm, clusters are laterally inhibited by each others’ activations. Laterally inhibited activations are considered to reflect the models’ overall familiarity with the current stimulus. The sum of these activations, Recognition Scores  $R$ , indexes this stimulus-specific familiarity. Lateral inhibition then ends in a winner-takes-all fashion — non-winning clusters’ activations will be muted for calculating further response probabilities.

Activations after lateral inhibition will spread to the category output units by weighted connections. The activations of each output units will then be turned into response probabilities. If the model made the correct response, then the winning cluster’s position is adjusted by moving it closer to the current input representation. In the event of a prediction error, an incorrect response, a new cluster centred on the current input representation is recruited. SUSTAIN prefers simple solutions, and only starts recruiting clusters in response to prediction errors. This means that more difficult tasks will cause SUSTAIN to densely populate the psychological space with clusters.

## Simulation of Edmunds et al. (2016)

In the following, we present a formal simulation with the SUSTAIN model capturing human categorization accuracy in II and UD structures. In addition, we show how the model captures better recognition memory following the II compared to the UD recognition task (Edmunds et al., 2016). This is supplemented by more clusters recruited for II, which mir-

rors the higher hippocampal activation while learning the II structures compared to UD structures (Carpenter et al., 2016). We do so by fitting SUSTAIN to an abstract design similar to that of Edmunds et al. (2016). We decided on Edmunds et al. (2016), because this allowed us to present the model with a close approximation of the conditions present where the authors observed better recognition performance in II.

Edmunds et al. (2016) used 36 grey squares that varied in brightness and size<sup>1</sup>. There were four conditions. UD structures included both horizontal and vertical category boundaries. II structures involved diagonal category boundaries with both positive and negative gradients. Figure 1 shows how these stimuli are distributed in the 2D physical stimuli space.

Each condition consisted of three phases. First, the training phase included 360 supervised training trials in blocks of 120. In each block, 24 stimuli randomly picked from the original 36 were shown 5 times. This was followed by an OLD/NEW recognition phase. This phase consisted of 3 blocks of the original 36 stimuli. The last phase was a categorization test phase. This phase was similarly made up of 3 blocks of the 36 original items. For a more detailed description of experimental procedure, see Edmunds et al. (2016). As SUSTAIN is a deterministic model, we only included a learning and a categorization test phase. After SUSTAIN stopped learning, its responses will remain the same. Recognition in SUSTAIN was assessed using Recognition Scores from categorization test phase, see details below.

## Model fitting

Our implementation of SUSTAIN is available in the R package *catlearn* (Wills et al., 2017, 2020). SUSTAIN’s parameters were adjusted to minimise the sum of squared errors between the overall mean categorization test phase accuracy of SUSTAIN and humans. The trial-order was randomised on each iterations with a random seed sampled in  $(0, 1000]$ . The model was fitted with a differential evolutionary algorithm, as implemented in the *DEoptim* package (Mullen, Ardia, Gil, Windover, & Cline, 2011). The algorithm reiterated 1000 times to find the best fitting parameters. The speed of crossover was set to  $c = 0.5$ , which gave larger weights to successful mutations. The top 30% best solutions were copied to the new iteration and was used in the new mutated population. These settings helped to find the single overall best parameter set for all category structures across different trial orders. The best fitting parameters with  $SSE = 3.8251454 \times 10^{-10}$  are presented in Table 1.

## Simulation

After finding the best fitting parameter across category structures, we simulated 1000 participants completing each four conditions using the parameters from Table 1. Each simu-

<sup>1</sup>In our simulations, these values were put in a range  $[0, 1]$  within each dimension. The values as specified by their respective coordinates are available in the supplementary material

Table 1: Best fitting parameters for SUSTAIN rounded to the 9<sup>th</sup> decimal place and their corresponding Upper Bounds set during the parameter search.

Parameters	Best Fitting
Attentional focus ( $r$ )	4.130121803
Lateral inhibition ( $\beta$ )	8.327280552
Decision consistency ( $d$ )	1.988279139
Learning rate ( $\eta$ )	0.062619803

lated participant received a seed in  $[1, 1000]$ . No two simulated participants received the same seed.

**Categorization Test Phase Accuracy** SUSTAIN’s categorization performance is qualitatively similar to what we observed from humans — II structures are harder to learn than UD. SUSTAIN matches human-level categorization test performance with a mean difference of 0.014, see Table 2.

Table 2: Categorization accuracy in SUSTAIN and humans.

Category Structures	SUSTAIN	Human
II	0.783	0.775
UD	0.848	0.868

**Cluster Recruitment and Attentional Tuning** The mean number of clusters recruited are  $M_{ii} = 5.59$ ,  $SD_{ii} = 1.20$  for II and  $M_{ud} = 3.01$ ,  $SD_{ud} = 1.18$  for UD. SUSTAIN solves II with a minimum of 4 clusters and a maximum of 12 clusters. SUSTAIN solves UD with a minimum of 2 and a maximum of 12 clusters. Example clusters populating the psychological space are shown in Figure 2.

The mean, and variation, in the number of clusters is the consequence of how trial-order interacts with the following mechanisms: similarity, attention and error-driven cluster recruitment. Simple problems on average result in fewer clusters, while harder problems require the recruitment of more clusters. This is attenuated by differentially weighing in relevant information from each dimension — by attentional tuning,  $\lambda$ . For example,  $\lambda$  is higher for relevant dimensions in UD structures, but remain comparable across dimensions in II, see Table 3.

Table 3: Mean  $\lambda$  values for each dimension across all category structures. UD = unidimensional; II = information-integration; +/- = vertical/horizontal for UD and positive/negative for II.

Stimuli Dimensions	II+	II-	UD+	UD-
X (Brightness)	13.52	13.47	13.38	7.80
Y (Size)	13.46	13.51	7.80	13.39

Overall, SUSTAIN requires higher number of clusters to

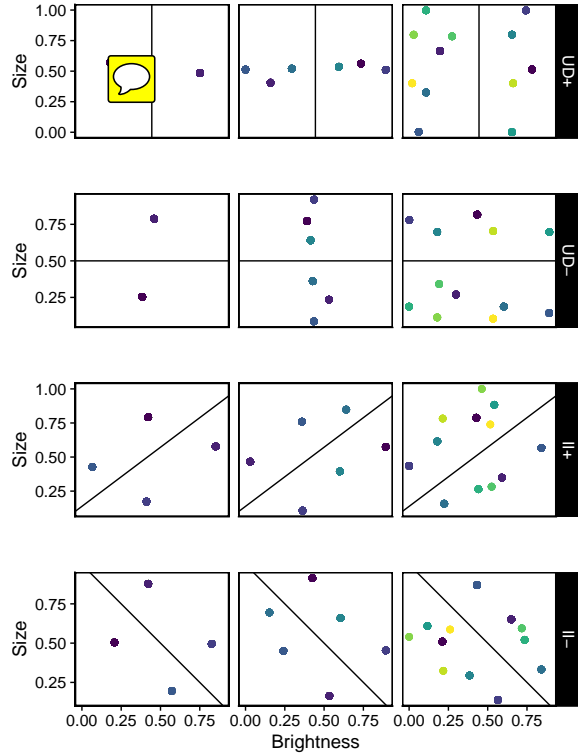


Figure 2: Example clusters recruited by SUSTAIN for three simulated participants across conditions. The juxtaposed black lines are the optimal decision bounds. UD = unidimensional; II = information-integration; +/- = vertical/horizontal for UD and positive/negative for II. Each colour is a separate cluster.

solve II due to its difficulty. This will result in clusters more perfectly matching training items, so the matching clusters will dominate the activation function. From the model’s point of view, we see higher HPC activations in II, because the category structure requires more representations to be encoded by the HPC. MTL has been shown to be responsible for similarity-based familiarity signals, and the HPC is specifically responsible for encoding new clusters after surprising events (Love & Gureckis, 2007). HPC activations have been shown to positively relate to cluster activations, updates and recruitments in SUSTAIN (Mack et al., 2016, 2018).

**Recognition** To get an approximate  $d'$  measure from  $R$ , Recognitions Score, we applied Equation A11 from Love and Gureckis (2007) to turn stimulus-specific  $R$  values during the categorization test phase into choice probabilities:  $P(old) = R / (R + k)$  where  $k$  is a response threshold parameter. We calculated the mean probability of a hit ( $P(H) = P(old | item_{old})$ ) and false alarm  $P(F) = P(old | item_{new})$  for each participant. We continued to determine  $d'$  for each participant using the z-transformed  $P(H)$  and  $P(F)$ . The we calculated group-level averages. This algorithm (including Equation 1 and the group-level  $d'$  calculations) were fitted against human performance in the recognition phase as indexed by  $d'$ . Simi-

larly, we used DEoptim and reiterated the parameter search 50 times. More details are included in the code available in the supplementary material. We found that the best-fitting parameter  $k$  was 0.571. This parameter will not change the ordinal pattern of the recognition performance ( $II > UD$ ) SUSTAIN shows given our previous fit to the categorization test data, but simply brings the values closer to human data. While this difference is rather small, it is statistically present in the between-groups comparison.

Table 4 shows the performance of humans and SUSTAIN. The model shows better recognition performance observed after learning II structures by Edmunds et al. (2016), but predicts higher false alarm rates for UD structures. A comparison of  $d'$  between SUSTAIN and human data yields a mean difference of 0.0071. This difference of  $d'$  and II and UD results from the difference in the number of recruited clusters between the two structures.

Recognition in SUSTAIN is based on similarity-driven cluster activation and lateral inhibition. Higher number of clusters recruited will be located more closely to the input representations encountered during test. This means that the stored representations will match better to the model’s previous experience in II than in UD problems. The more densely populated the psychological space with clusters, the more clusters neighbouring the input representation will activate. These activations then compete and will diminish as a result of lateral inhibition. The better recognition memory performance in II results from the higher sum of activations in regions neighbouring the input representations.

This benefit parallels HPC activation patterns. Better recognition memory performance follows not just from the modelling perspective, but also from a neural point-of-view. HPC has been long identified as crucial for memory (O’Reilly & Rudy, 2001; Schlichting & Preston, 2015). Love and Gureckis (2007) predicted this relationship, whereas high number of clusters mirror higher levels of HPC involvement. This prediction strongly aligns with Carpenter et al. (2016), who observed higher HPC involvement in the II compared to UD task, and our simulation, where SUSTAIN recruits more clusters for the II task.

Table 4: Mean recognition scores and  $d'$  for each category structure. Standard deviations are in parentheses.

	SUSTAIN $d'$	Human $d'$
II	0.040 (0.056)	0.01 (0.02)
UD	-0.016 (0.135)	0.00 (0.01)

## Discussion

Here, we have shown — for the first time — that both behavioural and neuroimaging data obtained in **these tasks** can be accommodated by a single-system model, SUSTAIN. The increased number of **category representations** for II also served as a base for better recognition memory performance



and larger HPC activation than UD because task-demands require a larger amount of information to be encoded in the HPC. Our simulation also extends prior fMRI work on the HPC's role in classification learning.

Previously, Davis et al. (2012) speculated that tasks like the II are not suitable to model with SUSTAIN. This sentiment was based on the idea that II is a procedural learning task ~~as it is learned procedurally~~ (Nomura et al., 2007) — characterized by mechanisms not specified within SUSTAIN. A procedural account (e.g. a simple weighted connectionist network) posits mechanisms that learn to associate visual inputs with behavioural (category) responses. However, procedural accounts of II problems are based on a range of experiments that received considerable scrutiny and often have alternative explanations (Newell, Dunn, & Kalish, 2011; Stanton & Nosofsky, 2007).

The only formal model — before SUSTAIN — that has been shown to accommodate both structures was COVIS. COVIS posited a procedural account of how people learn II structures. COVIS solves II with a procedural learning mechanism conceptualized as a three-layer network: the first layer calculates the exponent of the distance between activated input units and sensory units; the second layer attenuates these similarities by weighted connections between sensory units and striatal units before spreading to the striatal units; in the third layer, a decision rule responds with most activated striatal unit; and then the weights are updated. It is a distributed-representation connectionist network, where input node activations are supplied by a distance between sensory unit coordinates and input representation in the psychological space. COVIS solves UD by a different system, which establishes a decision bound dominating responding. Therefore, at limit COVIS predicts no recognition memory for either category structures. This still doesn't allow better recognition in II than UD. One approach would be to create an architecture that converts similarity derived from sensory input and weighted connections to activations of a memory traces. A similar approach has been used to describe recognition by multiple-traces memory models (Hintzman, 1986), but this can require the assumption that rule-based and procedural systems are able to access the where these values are stored.

## Conclusion

We formally show that a single-system adaptive clustering model, SUSTAIN, can accommodate categorization and recognition performance in two frequently used category structures, information-integration and unidimensional. The behaviour of the model is also consistent with MTL and HPC activity involved in learning these structures. Our simulation not only provides a formal account of how people learn these structures, but also contributes to the literature bridging formal models of category learning, behaviour and the brain.

## Open Science Statement

All simulation code is available in the Open Sciences Framework at [https://osf.io/jc9xs/?view\\_only=](https://osf.io/jc9xs/?view_only=)

c00913447a7d43deb1d61471c91b11ae.

## References

- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(1), 33.
- 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: Cambridge University Press.
- Ashby, F. G., & Valentin, V. V. (2018). The categorization experiment: Experimental design and data analysis. In E. J. Wagenmakers & J. T. Wixted (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, Fourth Edition, Volume Five: Methodology*. New York: Wiley.
- Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. N. (2016). A comparison of the neural correlates that underlie rule-based and information-integration category learning. *Human Brain Mapping*, 37, 3557–3574. doi: 10/gg58ps
- Casale, M. B., Roeder, J. L., & Ashby, F. G. (2012). Analogical transfer in perceptual categorization. *Memory & Cognition*, 40(3), 434–449. doi: 10/fmtxbh
- Davis, T., Love, B. C., & Preston, A. R. (2012). Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex*, 22(2), 260–273. doi: 10/c42v8p
- 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(4), 933–948. doi: 10/gg58pv
- 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(4), 840–859. doi: 10/gg58pt
- Edmunds, C. E. R., Inkster, A. B., Jones, P. M., Milton, F. N., & Wills, A. J. (2020). Absence of cross-modality analogical transfer in perceptual categorization: Keywords: Analogical reasoning, generalisation, strategies, learning, multiple systems. *Open Journal of Experimental Psychology and Neuroscience*, 3–13. doi: 10/ghbj3w
- Edmunds, C. E. R., Milton, F., & Wills, A. J. (2018). Due process in dual process: Model-recovery simulations of decision-bound strategy analysis in category learning. *Cognitive Science*, 1–28. doi: 10/gdqzc5
- Edmunds, C. E. R., Wills, A. J., & Milton, F. N. (2016). Memory for exemplars in category learning. In A. Papfragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2243–2248). Austin, TX: Cognitive Science Society.

- Edmunds, C. E. R., Wills, A. J., & Milton, F. N. (2019). Initial training with difficult items does not facilitate category learning. *Quarterly Journal of Experimental Psychology*, 72(2), 151–167. doi: 10/gg58pj
- Gureckis, T. M., & Love, B. (2004). Common mechanisms in infant and adult category learning. *Infancy*, 5(2), 173–198. doi: 10/dg6x54
- Gureckis, T. M., & Love, B. C. (2002). Who says models can only do what you tell them? Unsupervised category learning data, fits, and predictions. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th annual conference of the cognitive science society* (pp. 399–404). Hillsdale, NJ: Erlbaum.
- Gureckis, T. M., & Love, B. C. (2003). Towards a unified account of supervised and unsupervised category learning. *Journal of Experimental and Theoretical Artificial Intelligence*, 15(1), 1–24. doi: 10/fhzm7z
- Hintzman, D. L. (1986). "schema abstraction" in a multiple-trace memory model. *Psychological review*, 93(4), 411.
- 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(4), 503. doi: 10/gcx82s
- Le Pelley, M. E., Newell, B. R., & Nosofsky, R. M. (2019). Deferred feedback does not dissociate implicit and explicit category-learning systems: Commentary on Smith et al. (2014). *Psychological science*, 30(9), 1403–1409.
- Love, B. C. (2005). Environment and goals jointly direct category acquisition. *Current Directions in Psychological Science*, 14(4), 195–199. doi: 10/dfvrbg
- Love, B. C., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective & Behavioral Neuroscience*, 7(2), 90–108. doi: 10/ftn2cn
- Love, B. C., Markman, A. B., & Yamauchi, T. (2000). Modelling classification and inference learning. In H. Kautz & B. Porter (Eds.), *Proceedings of the seventeenth national conference on artificial intelligence* (pp. 136–141). Cambridge, MA: MIT Press.
- Love, B. C., & Medin, D. L. (1998a). Modeling item and category learning. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th annual conference of the cognitive science society* (pp. 639–644). Mahwah, NJ: Erlbaum.
- Love, B. C., & Medin, D. L. (1998b). SUSTAIN: A network model of human category learning. In C. Rich & J. Mostow (Eds.), *Proceedings of the fifteenth national conference on artificial intelligence* (pp. 671–676). Cambridge, MA: MIT Press.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111(2), 309–332. doi: 10/fm394p
- Mack, M. L., Love, B. C., & Preston, A. R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *Proceedings of the National Academy of Sciences*, 113(46), 13203–13208. doi: 10/f9fvzr
- Mack, M. L., Love, B. C., & Preston, A. R. (2018). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters*, 680, 31–38. doi: 10/gd5m53
- Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 204(1156), 301–328.
- Mok, R. M., & Love, B. C. (2019). A non-spatial account of place and grid cells based on clustering models of concept learning. *Nature Communications*, 10(1), 5685. doi: 10/ggftkf
- Mullen, K., Ardia, D., Gil, D., Windover, D., & Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6), 1–26. doi: 10/ggt446
- Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In *Psychology of learning and motivation* (Vol. 54, pp. 167–215). doi: 10.1016/B978-0-12-385527-5.00006-1
- Newell, B. R., Moore, C. P., Wills, A. J., & Milton, F. (2013). Reinstating the Frontal Lobes? Having More Time to Think Improves Implicit Perceptual Categorization: A Comment on Filoteo, Lauritzen, and Maddox (2010). *Psychological Science*, 24(3), 386–389. doi: 10/gg58pk
- Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., ... Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, 17(1), 37–43. doi: 10/dd3rdr
- Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition*, 33(7), 1256–1271. doi: 10/d3xpsz
- O'Reilly, R., & Rudy, J. (2001). Conjunctive Representations in Learning and Memory: Principles of Cortical and Hippocampal Function. *Psychological Review*, 108(2), 311–345.
- Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, 1, 1–8. doi: 10/gg775s
- Stanton, R. D., & Nosofsky, R. M. (2007). Feedback interference and dissociations of classification: Evidence against the multiple-learning-systems hypothesis. *Memory & Cognition*, 35(7), 1747–1758.
- Wills, A. J., Dome, L., Edmunds, C. E. R., Honke, G., Inkster, A., Schlegelmilch, R., & Spicer, S. (2020). *Catlearn: Formal psychological models of categorization and learning*.
- Wills, A. J., O'Connell, G., Edmunds, C. E. R., & Inkster, A. B. (2017). Progress in modeling through distributed collaboration. In *Psychology of learning and motivation* (pp. 79–115). Elsevier.