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

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Dome, Lenard
Edmunds, Charlotte
Wills, Andy J

Publication Date
2021

Peer reviewed

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

Lenard Dome (lenard.dome@plymouth.ac.uk)
School of Psychology, University of Plymouth
Plymouth, PL4 8AA UK

Charlotte E. R. Edmunds (ceredmunds@gmail.com)
Queen Mary, University of London
London, E1 4NS UK

Andy J. Wills (andy.wills@plymouth.ac.uk)
School of Psychology, University of Plymouth
Plymouth, PL4 8AA UK

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 participants learned these structures.

Keywords: categorization; recognition memory; formal model; SUSTAIN

Introduction

One commonly used pair of category structures in categorization research are the unidimensional (UD) and information-integration (II) category structures. UD and II structures were initially used for trying to separate the perceptual processes encoding the visual information from the 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 either vertical or horizontal decision bound: if the square is darker or larger than the set threshold, then it is category A, otherwise it is category B. Figure 1A and 1B shows that this optimal decision bound is parallel to one of the dimensional axes in the physical stimuli space. II structures are defined by diagonal optimal decision bounds. Figure 1C and 1D shows that II decision bounds follow a linear function, where the gradient is neither zero, nor infinite.

Many experiments utilised these structures (e.g. Carpenter, Wills, Benattayallah, & Milton, 2016; Donkin, Newell, Kalish, Dunn, & Nosofsky, 2015; Nomura et al., 2007; Le Pelley, Newell, & Nosofsky, 2019) and many initial empirical results were taken as evidence for COVIS (COmpetition between Verbal and Implicit Systems Ashby, Paul, & Maddox, 2011) — one formalization of a dual-system theory of categorization. Traditionally, dual-system theories have two distinct architectures using functionally different mechanisms. In COVIS, the explicit system uses rules that can be easily verbalized, while the implicit system maps perceptual input onto category responses. Accuracy in UD and II category structures, according to COVIS, depends on which system is engaged in solving the task. COVIS predicts that the explicit system will implement rules to optimally solve UD tasks, whereas the implicit system will take charge if simple rules are inadequate and implements (in this case) multidimensional strategies to combine information from the two dimensions of II tasks.

However, results from multiple labs pointed out flaws in the experimental designs in COVIS-inspired experiments (Newell, Moore, Wills, & Milton, 2013) with potential alternative explanations (Le Pelley et al., 2019; Donkin et al., 2015) or problems with the decision-bound analyses applied (Edmunds, Milton, & Wills, 2018; Edmunds, Wills, & Milton, 2019). In turn, some of these alternative-to-COVIS explanations have been critiqued (Ashby, Smith, & Rosedahl, 2019). The debate continues.

The current paper further examines some of the alternative-to-COVIS explanations of how people classifiy II and UD structures. Specifically, the way COVIS explains how people should optimally learn II structures was 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 when people were learning about II structures compared to when they were learning about UD structures. This result contradicts to predictions of COVIS as it is currently formalized, which posits less activation for II than UD. COVIS states that the explicit system is mapped to neurobiological substrates such as the MTL (HPC) and predominantly the prefrontal cortex, while the implicit system is mapeed to areas such as the supplementary motor areas and substantia nigra (Ashby & Valentin, 2017). According to COVIS, HPC and the prefrontal cortex is exclusively involved in the explicit

system, which is responsible for the optimal learning of UD structures. In other words, the way the two architectures are specified in COVIS are inconsistent with the differences in activations observed in HPC. HPC has also been long identified as crucial for memory (O'Reilly & Rudy, 2001; Schlichting & Preston, 2015). and thought to be essential for explicit memory. This suggest that people should have better recognition performance after learning II structures than in UD structures.

Given Carpenter et al. (2016)'s observation of greater HPC activity in II than in UD structures, one can further predict, contrary to COVIS, that there will be better post-recognition memory for exemplars in II than in UD structures. Edmunds et al. (2016) directly confirmed this prediction. They found better recognition memory after learning II than UD structures, essentially supplementing the neural data. While the differences in recognition performance are rather small, it is statistically present in a between-groups comparison (Edmunds & Wills, 2016). A more extensive investigation on recognition memory in UD and II problems also found that participants who reported using complex multidimensional rules showed better recognition performance (Edmunds, 2017).

Building on these findings, we further supplement behavioral and neural data with evidence from computational modeling. Here, we provide a formal single-system explanation of the results of both Carpenter et al. (2016) and Edmunds et al. (2016). We do so by using SUSTAIN (Supervised and Unsupervised STratified Adaptive Incremental Network Love et al., 2004).

SUSTAIN is one model in a single-system approach to modeling categorization, and is able to accommodate a wide range of behavioral and neural phenomena (e.g. Love et al., 2004; Gureckis & Love, 2004; Davis, Love, & Preston, 2012). 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 can 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, has been mapped to HPC and MTL functions and activations.

Cluster-specific model components in SUSTAIN have been directly connected to strong HPC activations present in early learning and low HPC functions in amnesic patients in a dot-pattern classification task (for a more exhaustive review, see 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 modeling to neural activity patterns. For example, during rule-plus-exception learning, SUSTAIN makes specific predictions about item recognition, which has been 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 accommodates behavioral responses and HPC activity in subsequent learning tasks where the stimuli remain perceptually the same, but irrelevant features in the first task become essential in the new categorization problem (Mack, Love, & Preston, 2016). 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).

More difficult tasks require SUSTAIN to bind (and store) larger sets of information into clusters than simpler tasks do (Love et al., 2004). This process results in higher number of clusters being recruited, which has been previously mapped to increased HPC activity and improved recognition accuracy (Love & Gureckis, 2007). SUSTAIN therefore predicts better recognition performance in tasks that require higher number of clusters. In this paper, we intend to test these predictions in relation to UD and II category structures by formally fitting SUSTAIN to the categorization accuracy data of Edmunds et al. (2016). Furthermore, we will compare its recognition performance to human recognition performance from Edmunds et al. (2016) and evaluate whether the cluster-recruitment process is consistent with increased HPC activity in II problems compared to UD problems observed in Carpenter et al. (2016).

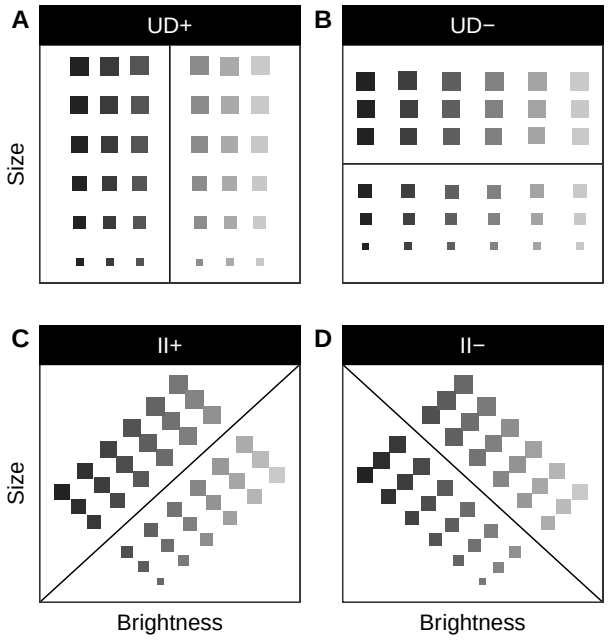

Figure 1: Representations of the category structures used in Edmunds et al. (2016). UD = unidimensional; II = information-integration; +/- = vertical/horizontal for UD and positive/negative for II.

We refer the reader to Love et al. (2004) for the full de-

scription 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, which proposes that clusters underlie category representations (Love et al., 2004). Clusters, from SUSTAIN's perspective, are single coordinates in the representation space. These coordinates are internal representations that connect to categories. SUSTAIN starts with one cluster, centered 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 score *R*, indexes this stimulus-specific familiarity. Lateral inhibition then ends in a winner-takes-all fashion — non-winning clusters' activations are muted for calculating further response probabilities.

Activations after lateral inhibition spread to the category output units by weighted connections. The activations of each output units are 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 centered on the current input representation is recruited and becomes the winning cluster. Connection weights from cluster units to the category output units are updated according to the one-layer delta learning rule (Widrow & Hoff, 1960). After both correct and erroneous responses, the winning cluster updates SUSTAIN's attentional tuning. Attentional tuning of each dimension maximizes its impact on the recruited clusters. 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 accommodating human categorization accuracy in II and UD structures. In addition, we show how the model captures categorization accuracy and predicts better recognition memory following the II problems compared to the UD problems (Edmunds et al., 2016). This difference should be based on more clusters recruited for II, which leads to the prediction of 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 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[1]. There were four conditions. UD structures included both vertical and horizontal category boundaries, shown on Figure 1A and Figure 1B respectively. II structures involved diagonal category boundaries with both positive and negative gradients, shown on Figure 1C and Figure 1D respectively.

Each condition consisted of three phases. First, the categorization training phase included 360 supervised training trials in blocks of 120. Each simulated participant received 24 stimuli randomly picked from the 36 for their simulation. Those 24 stimuli were shown 5 times in each of the 3 blocks. This was followed by an OLD/NEW recognition phase. This phase consisted of 3 blocks of all 36 stimuli. The last phase was a categorization test phase. This phase was similarly made up of 3 blocks of all 36 items. For a more detailed description of experimental procedure, see Edmunds et al. (2016).

Simulation

Our implementation of SUSTAIN is available in the R package catlearn (Wills et al., 2020). First, we wanted to find the one best fitting parameter set for the model across all four categorization problems. SUSTAIN therefore encountered all four problems at the same time as a single participant - SUSTAIN completed each problem once with the same set of parameters. SUSTAIN was reset between each problem. SUSTAIN's parameters were then adjusted to minimise the sum of squared errors (SSE). SSE was calculated between the mean group-level accuracy of humans in the categorization test phase (as reported in Edmunds et al. (2006) and shown in Table 2) and the mean accuracy of SUSTAIN during the categorization test phase. We used the group-level data, because it captures the ordinal difference associated with these tasks: participants show higher accuracy for UD than II. The trial order was randomised on each iteration. The model was fitted with a differential evolutionary algorithm, as implemented in the DEoptim package (Mullen, Ardia, Gil, Windover, & Cline, 2011). The algorithm iterated 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 set of parameters for all category structures across different trial orders. The best fitting parameters are presented in Table 1. After finding the best set of parameter, we simulated 1000 different trial orders with SUSTAIN.

[1] 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 4th decimal place.

Parameters	Best Fitting
Attentional focus (r)	4.1301
Lateral inhibition (β)	8.3273
Decision consistency (d)	1.9883
Learning rate (η)	0.0626

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 in Edmunds et al. (2016). This difference in accuracy is a reliable difference in SUSTAIN's performance, $BF = 1.88 \times 10^{776}$. 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. Standard deviations are in parenthesis.

Category Structures	SUSTAIN	Human
II	0.78 (0.027)	0.78 (0.11)
UD	0.85 (0.026)	0.87 (0.07)

Cluster Recruitment and Attentional Tuning The mean number of clusters recruited were $M_{ii} = 5.59$, $SD_{ii} = 1.20$ 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 of perceptual inputs. Each dimension has its own attentional tuning, λ. For example, λ is higher for relevant dimensions in UD structures, but remains comparable across dimensions in II, see Table 3.

Table 3: Mean λ values for each dimension across all category structures. UD = unidimensional; II = information-integration; +/- = vertical/horizontal for UD and positive/negative for II. Standard deviations are in parenthesis.

Conditions	λ_x	λ_y
II+	13.52 (0.81)	13.46 (0.78)
II-	13.47 (0.78)	13.51 (0.81)
UD+	13.38 (1.48)	7.80 (1.41)
UD-	7.80 (1.41)	13.39 (1.48)

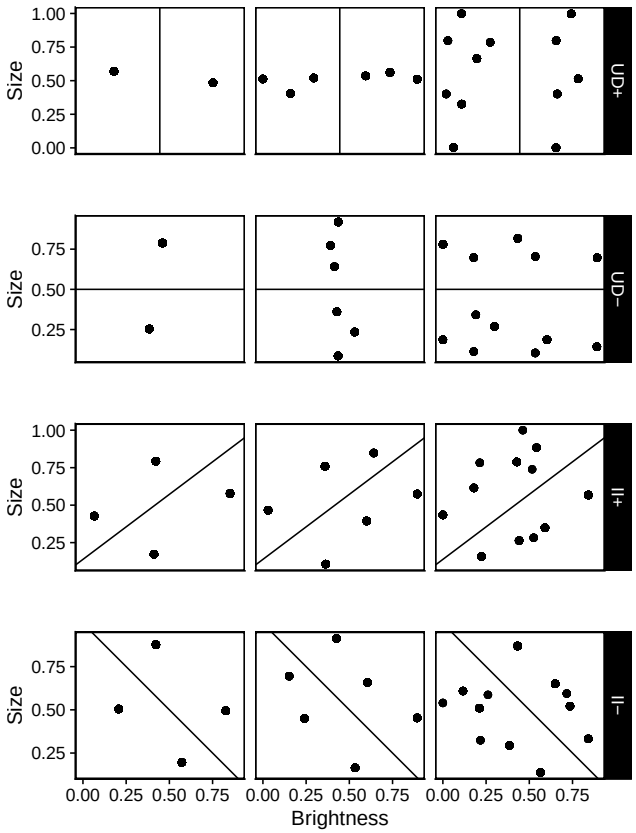

Figure 2: Example clusters recruited by SUSTAIN for three simulations across conditions. The juxtaposed black lines are the optimal decision bounds. UD = unidimensional; II = information-integration; +/- = vertical/horizontal for UD and positive/negative for II.

If SUSTAIN tries to incorporate the irrelevant dimension in UD by attending to both dimensions equally, the only way SUSTAIN can eventually solve the task is to recruit more clusters. Similarly, if SUSTAIN only attends to a single dimension in II, it will need to recruit large number of clusters to solve the task. This will also result in misclassification during the categorization test phase. Figure 3 shows how the 36 stimuli are captured by different clusters (indicated by the dots' color) during the categorization test phase. Figure 3 second row gives an example when clusters from one side of the optimal decision bound captures stimuli from the other side of the decision bound. This is due to a single dimension weighting in more in cluster activations than the other dimension.

Overall, SUSTAIN requires higher numbers of clusters to 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, the HPC is specifically responsible for encoding new clusters after surprising events (Love & Gureckis, 2007). We see higher HPC activations in II, because the category structure requires more representations to be encoded by the HPC.

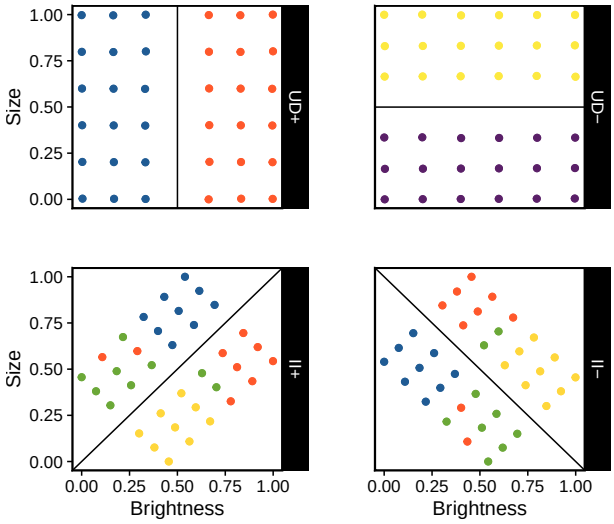

Figure 3: Example physical stimuli space for a single simulation during the categorization phase. Each color is a different cluster SUSTAIN recruited during learning. The dots with same colors are captured by the same cluster. The juxtaposed black lines are the optimal decision bounds. UD = unidimensional; II = information-integration; +/- = vertical/horizontal for UD and positive/negative for II.

HPC activations have been shown to positively relate to cluster activations, updates and recruitments in SUSTAIN (Mack et al., 2016, 2018). Therefore, SUSTAIN's prediction for the difference in HPC activity between UD and II problems is found to be consistent with Carpenter et al. (2016).

Table 4: Mean recognition scores and d' for each category structure, rounded to three decimal places. Standard deviations are in parenthesis.

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

Recognition To get an approximate d' measure from R, Recognition 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 $(P(H) = P(old \mid item_{old})$ and false alarm $P(F) = P(old \mid item_{new})$ for each simulated participant. We continued to determine d' for each participant using the z-transformed $P(H)$ and $P(F)$. Then we calculated group-level averages. This algorithm (including the group-level d' calculations) were fitted against human performance in the recognition phase as indexed by d'. Similarly, we used DEoptim and reitereted the paramater 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 the simulated categorization test data, but simply brings the values closer to the human data.

Table 4 shows the performance of humans and SUSTAIN. A comparison of d' between SUSTAIN and human data yields a mean difference of 0.023. The model predicts better recognition performance after learning II than UD structures, consistent with Edmunds et al. (2016). This is a realiable difference in the simulated data, BF = 7.06×10^{57}. This difference of d' between SUSTAIN's recognition performance in 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. Where SUSTAIN recruits a large number of clusters, these clusters will generally be closer to the stimulus representations presented in the recognition memory test. This means that the stored represention 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 modeling perspective, but also from a neural point-of-view. Love and Gureckis (2007) predicted this relationship, where higher 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.

Discussion

We have presented a formal account of empirical results (Edmunds et al., 2016; Carpenter et al., 2016) concerning the acquisition of unidimensional (UD) and information-integration (II) category structures. In so doing, we have shown - for the first time - that both the behavioral and neuroimaging data obtained in these tasks can be accomodated by a single-system model, SUSTAIN. The increased number of clusters recruited by SUSTAIN for the II structure served as a base for better recognition memory performance, and larger HPC activation, than in the UD structure. According to SUSTAIN, this is because the differing task demands of the two structures requires a larger amount of information to be encoded in the HPC for II structures.

Previously, Davis et al. (2012) speculated that tasks like the II category learning were not suitable to model with SUSTAIN. This sentiment was based on the idea that II category learning is a procedural learning task (Nomura et al., 2007) —

and hence characterized by mechanisms not specified within SUSTAIN. However, procedural accounts of II problems are based on a range of experiments that received considerable scrutiny, and which turn out to have alternative explanations.

While the findings reported here are preliminary, they provide a sufficient explanation for a range of findings related the UD and II structures in the form of a fully specified formal computational model — SUSTAIN. Nonetheless, our current simulation might be considered unconstrained, because we did not pursue quantiative fit per se. Instead we choose to focus on whether SUSTAIN could accomodate the UD/II differences in performance during the categorization test. We then investigated the predictions SUSTAIN made on that basis relating to the subsequent recognition task, and the differences in MTL/HPC activations across UD and II. One promising follow-up would be to explore individual differences in HPC activations and categorization accuracy via fitting SUSTAIN to subject-level results. This would allow a direct mapping between cluster recruitment and HPC activations. A caveat with this approach is SUSTAIN's sensitivity to trial-order effects.

The only formal model — before SUSTAIN — that has been argued to accomodate the classification of both UD and II 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 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 the 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, rule-based 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 memory traces. A similar approach has been used to describe recognition by multiple-trace memory models (Hintzman, 1986), but this can require the assumption that both rule-based and procedural systems are able to access the representation space 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 behavior 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, behavior and the brain.

Open Science Statement

All simulation code is available in the Open Sciences Framework at `https://osf.io/jc9xs/`.

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., Smith, J. D., & Rosedahl, L. A. (2019). Dissociations between rule-based and information-integration categorization are not caused by differences in task difficulty. *Memory & cognition*, 1–12.

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 (second edition)* (Second Edition ed., p. 157-188). San Diego: Elsevier.

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.

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.

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.

Edmunds, C. E. R. (2017). *Critique of a dual-system model of category learning*. Unpublished doctoral dissertation, University of Plymouth.

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.

Edmunds, C. E. R., & Wills, A. J. (2016). Modeling category learning using a dual-system approach: A simulation of Shepard, Hovland and Jenkins (1961) by COVIS. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, 69–74.

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.

Gureckis, T. M., & Love, B. (2004). Common mechanisms in infant and adult category learning. *Infancy*, *5*(2), 173–198.

Hintzman, D. L. (1986). " schema abstraction" in a multiple-trace memory model. *Psychological review*, *93*(4), 411.

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., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective & Behavioral Neuroscience*, *7*(2), 90–108.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332.

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.

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.

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.

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.

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.

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.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. In *Institute of radio engineers, western electronic show and convention, convention record* (Vol. 4, p. 96–104).

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.* (R package version 0.7.5)

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.