Title
Categorization in Environments that Change when People Learn

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Villarreal, Manuel
Vaday, Sahar
Lee, Michael

Publication Date
2022

Peer reviewed

Categorization in Environments that Change when People Learn

Manuel Villarreal (jesusmv1@uci.edu)
Sahar Vaday (svaday@uci.edu)
Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences, University of California, Irvine
3151 SSPA, Irvine, CA 92697-5100 USA

Abstract

Most studies of human category learning involve category structures that do not change, or that change in a way that is independent of people's categorization behavior. We consider the situation in which successful category learning causes categories to change. In an experiment, participants learned from feedback whether animals are healthy or diseased. Once their categorization accuracy was near-perfect, the category structure changed so that different animals became diseased. Based on exploratory data analysis and the application of two category learning models, we argue that, once they detect a category change, people retain what they have learned about healthy animals, but reset what they have learned about diseased animals. We discuss future modeling goals and emphasize the need for learning models to study situations in which people's behavior impacts the dynamics of the environment in which learning takes place.

Keywords: category learning; changing environments; dynamic environments; categorization models

Introduction

Most studies of human category learning involve fixed categories (e.g., Feldman, 2000; Shepard, Hovland, & Jenkins, 1961; Smith & Minda, 2000). This is appropriate for understanding how people learn about stable concepts. It is reasonable to assume that many natural kinds —fruits, insects, weapons, and so on—have stable relationships between stimuli and categories. For example, the assignment of rocks to categories like obsidian, basalt, and granite involves a stable category structure (Nosofsky, Sanders, & McDaniel, 2018). The assignment of colors to categories like red, blue, and yellow involves cultural differences in the available categories and assignments (Regier, Kay, & Khetarpal, 2007), but those structures are largely stable within a culture.

Some category learning studies use more dynamic environments, in which categories change over time. The change could be a sudden reassignment of stimuli to different categories, or a gradual drift in the probability that stimuli belong to categories (e.g. Estes, 1984; Gallistel, Mark, King, & Latham, 2001; Navarro, Perfors, & Vong, 2013; Speekenbrink & Shanks, 2010; Kruschke, 1996). These tasks are appropriate for understanding how people adapt to new category structures and non-stationary environments. Most of these previous studies determine the dynamics of environmental change ahead of time, and assume that change is independent of participant behavior. Category learning studies rarely consider dynamic environments that change *in response to* people's decisions.[1] Assuming that category learning is independent of category structure may be appropriate in some situations, at least as an approximation. For example, starting from house telephones, the technological development that led to the sudden introduction of car phones, then mobile phones, and then smartphones has required people to change how they categorize stimuli as phones. This learning process, however, has not influenced technological development. As another example, the seasons drift cyclically largely independent of the categories people learn. This means that people adapting their categorization from Finland being an undesirable vacation destination in winter to a desirable one in summer does not influence the weather in Helsinki.

These examples hint, however, at the limits of the independence assumption. People's ability to learn to use new devices as phones creates longer-term markets for technological development. Similarly, the independence of people's behavior from temperature fluctuation only holds for those categories and time scales that do not involve human-influenced global warming. It is generally not the case that the dynamics of an environment are completely decouple from people's learning and behavior in that environment.

Accordingly, it is not hard to identify real-world category learning situations in which people's learning and environmental dynamics are tightly coupled, with changes in categorization behavior leading to changes in the category structures being learned. In the natural world, one cause of virus mutation, along with copying error and select cell pressure, is a change in the immunity of potential hosts (Sugak, Martynyuk, & Drozd, 2015). This means that as society develops better treatments the virus environment changes. Loosely speaking, as the categorization problems involved in providing immunity—correctly classifying treatments as effective or not effective—is solved, the categorization problem itself changes as a consequence. In the human-constructed world, phishing scams continually need to adapt to evade spam filters (El Kouari, Benaboud, & Lazaar, 2020). The spam filters

[1]Perhaps the closest example is the Wisconsin Card Sorting Test (WCST: Dehaene & Changeux, 1991; D'Alessandro, Radev, Voss, & Lombardi, 2020), in which participants have to organize a set of cards base on an underlying rule. As performance improves the rule can change. The main difference is that rules in the WCST are typically based on a single stimulus dimension, such as color or shape. In general, the relationship between stimuli and categories is more complicated than a single dimension.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

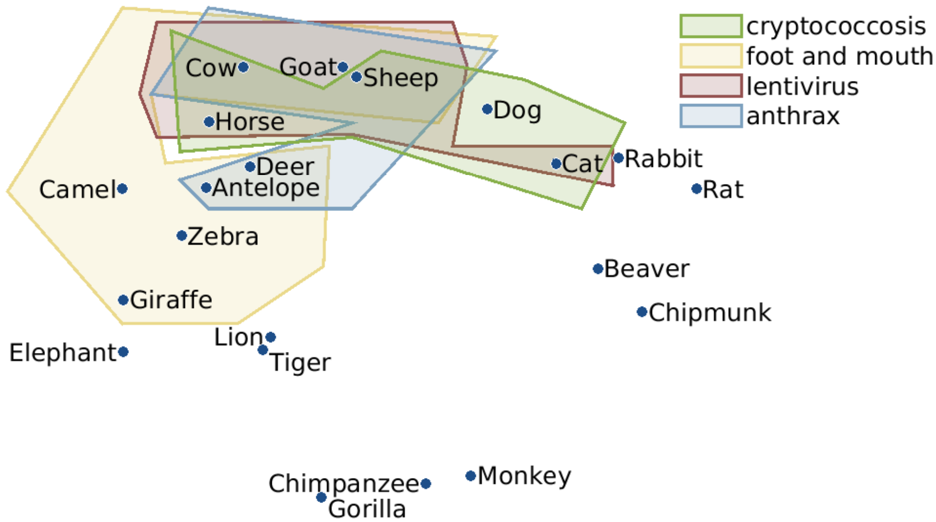

Figure 1: The four category structures. The 21 animal stimuli are represented as points arranged so that more semantically similar animals are located nearer each other. The four diseases are represented by colored regions that encompass those animals that have the disease.

solve a categorization problem to separate email stimuli into legitimate and blocked categories. As the accuracy of filters improves, the nature of the categorization problem changes, with new phishing attacks developed.

In this paper we study people's category learning behavior in a task for which the category structure being learned changes when people become sufficiently accurate. Over a sequence of trials, people are asked to categorize animals as healthy or infected with some disease, based on feedback provided after every trial. Once they reach a high level of accuracy, the category structure changes, so that a different set of animals become diseased. The environmental change is not signaled other than through the change in feedback for specific animals on individual trials. We are interested in how people perform in this learning situation, for which environmental dynamics are linked to their category learning.

Experiment

Participants

38 undergraduate student participants at the **redacted to satisfy anonymous submission requirements** completed the experiment for course credit.

Stimuli

The stimuli were 21 animals and the four category structures corresponded to four real-world diseases: cryptococcosis, foot and mouth, lentivirus, and anthrax. Figure 1 shows the set of stimuli, and their assignment to the healthy and diseased categories for all four diseases. The animals are represented as points using non-metric multidimensional scaling as a visualization method (Kruskal, 1964), based on similarity data reported by Westfall and Lee (2021). The animals with each of the diseases are contained within colored regions.

It is clear there is considerable overlap between the four category structures and the differences between them are subtle. Sheep and cow belong to the diseased category for all of the diseases, horse and deer belong to the diseased category for exactly half the diseases, and a large number of animals always belong to the healthy category. The category structures vary between five and eight diseased animals, so disease is always the lower base-rate category.

Procedure

All participants completed 210 categorization trials. On each trial, an animal was presented as a picture with an accompanying text label. The same picture was used every time that animal was presented. Participants were required to categorize the animal as "healthy" or "diseased". They then received feedback of the form "wrong, the horse is diseased", informing them whether their response was correct and making explicit the correct classification. At the top of the interface a set of 210 slots was shown, corresponding to the 210 trials. Completed correct responses were shown as black circles, completed incorrect responses were shown as crosses, and trials yet to be completed were shown as gray circles. This information was updated after every trial.

If, at any point in the sequence of trials, the participant had correctly categorized 18 or more out of the last 21 animals, the category changed. The study information sheet told participants that "It is possible that whether or not a particular animal is healthy could change over the course of the experiment," but a change in category structure was not indicated in any way during the experiment. The animals were presented in a random order, subject to the constraint that no animal be presented twice within the same disease category until all other animals had been presented. The experiment was completed after 210 trials, regardless of the participant's accuracy.

Three different sequences of transitions from one category structure to the next were used. We refer to these sequences as conditions. Condition 1 started with anthrax, followed by lentivirus, cryptococcosis, and foot and mouth. Condition 2 started with anthrax, followed by foot and mouth, cryptococ-

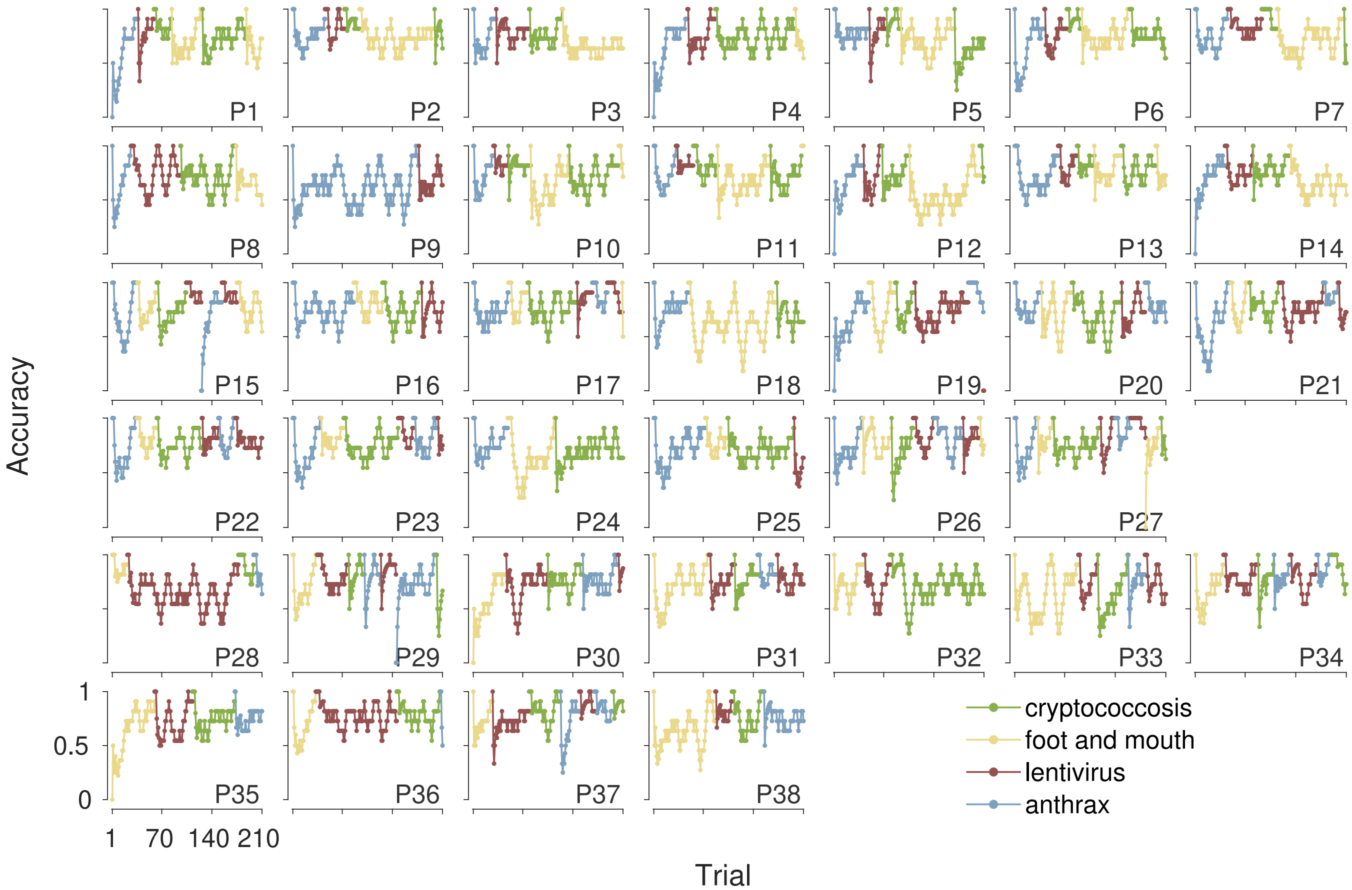

Figure 2: Category learning performance for all 38 participants. Each panel corresponds to a participant, with colored lines showing their average proportion of correct responses on the last 10 trials with the current category. Changes in category are shown by different colors. The participant panels are arranged so that the top two rows correspond to the first condition, the middle two rows correspond to the second condition, and the bottom two rows correspond to the third condition.

cosis, and lentivirus. Condition 3 started with foot and mouth, followed by lentivirus, cryptococcosis, and anthrax. A total of 14, 13, and 11 participants completed conditions 1, 2, and 3 respectively, in a between-participants design. These sequences were intended to allow comparisons that focus on specific research questions. For example, always having cryptococcosis as the third disease allows a controlled comparison of the impact of the previous two diseases on learning.

Results

Figure 2 shows the category learning performance of all 38 participants. Each panel corresponds to a participant and the panels are organized by condition. The colored lines show the average proportion of correct responses over the last 10 trials for the current disease category. Different diseases are indicated by different colors. For most participants, there is a clear pattern of a sudden decrease in accuracy following a change and then subsequent learning of the new disease category. After learning the first disease in their sequence, most participants maintain an average accuracy well above chance for the remaining diseases, which suggests some beneficial transfer of learning from one category to the next.

There are also clear individual differences. For example, participant 28 learns the first foot and mouth disease category quickly, whereas participant 38 takes many trials to reach the criterion level of accuracy. Interestingly, however, participant 28 then takes many trials to learn the subsequent lentivirus disease, whereas participant 38 now learns quickly.

Trials Needed to Learn Categories

Each of the four disease categories are about equally difficult to learn. Aggregated over all participants, and all of their attempts at learning the categories, the mean (standard deviation) number of trials to learn is 46.7 (22.0) for cryptococcosis, 52.0 (27.5) for foot and mouth, 39.0 (26.8) for lentivirus, and 41.8 (25.2) for anthrax. The Bayes factor for a one-way ANOVA is greater than 1000 in favor of these distributions having the same mean, rather than independently different means. This is evidence that the average number of trials that it takes participants to learn does not depend on the category.

Figure 3 provides a vertical histogram of learning times, considering both the disease category and its position in the

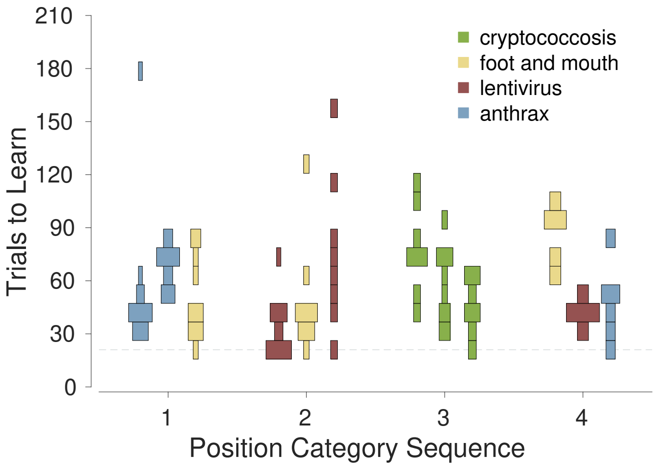

Figure 3: The distribution of the number of trials needed to learn the category at each position in the sequence. The distributions for conditions 1, 2, 3 are shown from left to right at each position. Distributions are colored according to the disease category. The dotted gray line shows the minimum of 21 trials needed to demonstrate learning.

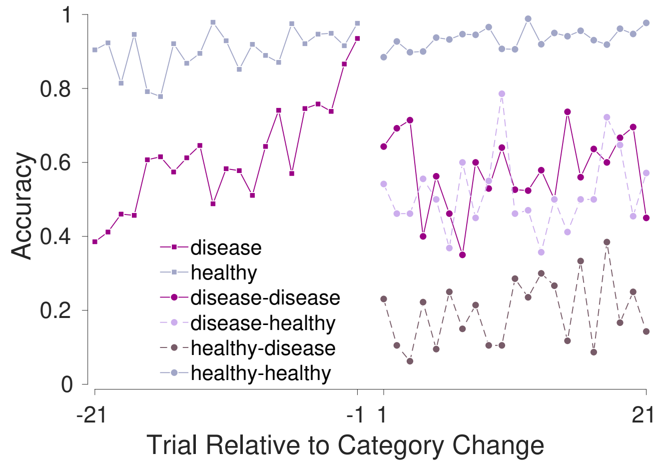

Figure 4: Accuracy for different patterns of change between diseased and healthy animals across category changes. The four lines correspond to the different transitions between healthy and disease categories, showing the average accuracy across all participants, stimuli, and changes for the 21 trials leading up to the change and the 21 trials after the change.

learning sequence. The width of each square shows the frequency with which participants took that number of trials to learn that category in that position . Only the first four positions are considered, corresponding to the first time a participant encountered each disease category. For each position, there are three possibilities, corresponding to the three conditions. There is little evidence of differences in the learning distributions across the four positions. The Bayes factor for a one-way ANOVA is greater than 1000 in favor of sameness. Comparing the same disease in different positions also shows few differences. A t-test comparison of group means for anthrax in the first versus fourth positions provides an inconclusive Bayes factor of 2.0 in favor of a difference. Comparing lentivirus in the second and fourth positions provides a Bayes factor of 5.4 in favor of sameness. Comparing the three distributions of cryptococcosis which involve different prior learning experiences across the conditions, a one-way ANOVA provides a Bayes factor of 9.3 in favor of sameness.

Overall, there are neither strong nor systematic differences in the distributions of the number of trials needed to learn the different categories at different positions in the sequence. This is an interesting finding. On the one hand, the overlap between the different categories shown in Figure 1 means there is clearly some transfer advantage from prior learning. Many of the animals learned to be healthy, for example, will remain healthy. On the other hand, the similarity of the categories means prior learning could interfere with the fine-grained distinctions needed to master a new disease. The results in Figure 2 suggest these transfer and interference effects tend to balance each other out.

Accuracy Before and After Category Changes

There are four possible patterns of category association for an animal over a change in category structure. An animal can be diseased in both categories, healthy in both, change from being healthy to diseased, or change from being diseased to healthy. Figure 4 shows the change in accuracy for these four different possibilities, for the 21 trials following a category change. It also shows accuracy for the diseased and healthy animals in the original category structure for the 21 trials leading up to a category change. The measures of accuracy are aggregated over all participants, animal stimuli, and disease category transitions.

Leading into the category change, most learning is evident for the diseased animals, with the healthy animals generally already being accurately categorized throughout. After the category change, those animals that remain healthy continue to be accurately categorized. Animals that continue to be diseased, in contrast, are suddenly much less well categorized, with accuracy falling to around 50%. This is about the same level as animals that have changed from being healthy to diseased. There is an even more drastic drop in accuracy for animals that were healthy but have become diseased.

One interpretation of this pattern of results is that participants assume that the healthy animals continue to be healthy after a category change, but decide to re-learn the diseased animals. The assumption of stability in healthy animals is consistent with high healthy-healthy accuracy but low healthy-disease accuracy. The assumption of re-learning the diseased category is consistent with disease-disease and disease-healthy having the same moderate accuracy, independent of whether or not the now diseased animals changed category.

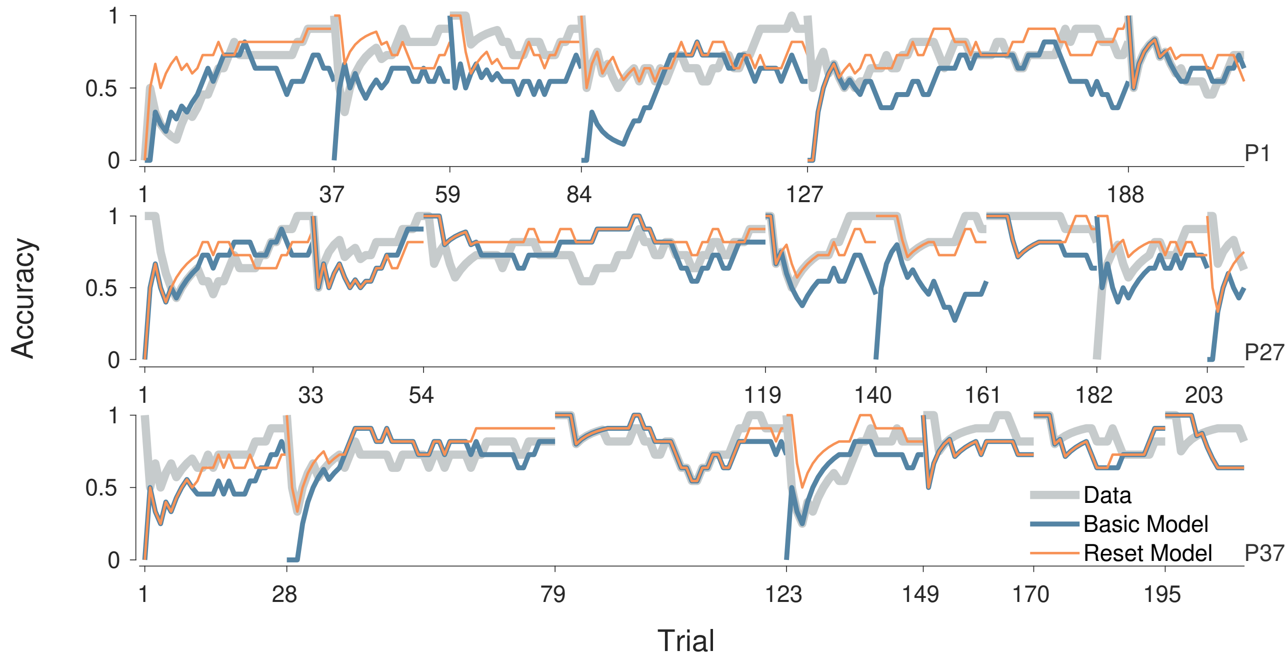

Figure 5: Model performance for three representative participants. The lines show the smoothed accuracy of the behavioral data, the basic model, and the reset model. The x-axis tick marks indicate the trials at which a category change occurred.

Modeling

Many standard models of trial-by-trial category learning with feedback rely on incremental learning rules that adjust the associations between stimuli and categories (Kruschke, 2008; Shanks, 1991). Combining this approach to learning with similarity-based generalization gradients, based on exemplar representation of stimuli, has been shown to avoid catastrophic forgetting (Kruschke, 1993), and leads to the influential ALCOVE model (Kruschke, 1992) and its variants. We base our modeling on the version of ALCOVE developed by Lee and Navarro (2002) that relies on feature-based representations, since the animal stimuli seem better represented in terms of high-level cognitive features than low-level perceptual dimensions.

Our empirical results, especially through the analysis in Figure 4 suggest that adaptation to category change can be understood in terms of what associations are preserved and reset when the category changes. To explore this intuition, we compare a basic model with only incremental learning against an extended model that resets the associations for previously diseased animals after each category change.

Two Learning Models

Formally, the similarity between animal i and j is represented by s_{ij} which are calculated as

$$s_{ij} = \exp\left(-\sigma\left[\sum_x f_{ix}(1-f_{jx}) + \sum_x (1-f_{ix})f_{jx}\right]\right), \quad (1)$$

where the f_{ix} are binary features, with $f_{ix} = 1$ if the animal has feature x and $f_{ix} = 0$ if it does not. The combination of features in Equation 1 provides a measure of the difference between animals i and j, consistent with the contrast model (Tversky, 1977). The exponentiation corresponds to a standard form of generalization gradient (Shepard, 1987), with a decay parameter $\sigma > 0$. We use the features for the animal stimuli found using similarity modeling by Westfall and Lee (2021). The association between the animal i and category k (i.e., healthy or diseased) is represented by a weight w_{ik}, all of which start at zero for the first category. When animal i is presented, the overall response strength for each category is calculated as

$$r_{ik} = \sum_j s_{ij} w_{jk}, \quad (2)$$

which provides a response probability

$$P(R = k \mid i) = \frac{\exp(\phi r_{ik})}{\sum_g \exp(\phi r_{ig})}, \quad (3)$$

where $\phi > 0$ is a response determinism parameter.

Once a decision has been made and feedback received, a standard delta learning rule is used to update the association weights (Sutton & Barto, 1998)

$$\Delta w_{jk} = \lambda(t_k - r_{ik}) s_{ij}, \quad (4)$$

where $0 \leq \lambda \leq 1$ is a learning rate parameter and t_k is the teacher signal for category C_k. Following Kruschke (1992) the "humble" teacher feedback is

$$t_k = \begin{cases} \max(+1, r_{ik}) & \text{if } i \in C_k \\ \min(-1, r_{ik}) & \text{if } i \notin C_k. \end{cases} \quad (5)$$

The learning rule for the weights is designed to minimize the sum-squared difference between the response strengths and teacher values.

In the extended "reset" model, associations for the disease category are reset to zero when a category changes. This is consistent with the observation that accuracy for animals that had previously been diseased is relatively low after the category change, regardless of their new category membership.

Modeling Results

We applied both the basic and reset models to the category decisions made by the 34 participants who performed well enough to encounter each disease category at least once. The two models were fit independently for each participant using maximum likelihood, optimizing the σ, ϕ, and λ parameters. At the maximum-likelihood values, the basic model agrees with the behavioral data for 76% of trials. The reset model agrees on 81% of trials. This improvement is consistently shown at the participant level, with 30 out of 34 participants better described by the reset model.

Figure 5 provides insight into how the reset model improves upon the basic model. It shows the accuracy over trials of both models and the behavioral data for three representative participants, with one participant chosen from each condition. The reset model is generally able to describe performance between category changes a little better than the basic model, although both are far from perfect. The reset model is often significantly better, however, at describing the participants' behavior immediately after a category change. The basic model regularly shows very low accuracy for a few trials, whereas the reset model shows patterns of accuracy qualitatively more consistent with participant performance. This discrepancy does not happen after every category change. There are exceptions in which a participant does drop to very low accuracy after a category change. Overall, however, the additional assumptions in the reset model seem to capture an important aspect of participant behavior that often occurs.

Limitations and Extensions

There is an obvious need to test the generality of our results using other stimuli, categories, and category structures. In particular, as Figure 1 shows, all of the disease categories we used had large overlap, and contained a minority of stimuli. Other base-rates and greater variability between categories need to be considered.

Our modeling also provides only a small first step toward understanding people's behavior. The reset mechanism is crude and there are plausible alternatives. In particular, attention shifts provide a mechanism that could account for how people quickly learn as categories change (Kruschke, 2003). In addition, the reset model assumes that people detect the category change accurately and immediately. A more complete model needs to account for how people identify that the category structure has changed. There are some psychological theories and cognitive models of adaptation and self-regulation in learning that provide possible starting points (e.g., Lee, Newell, & Vandekerckhove, 2015).

There is also a broader cognitive science literature from artificial intelligence and machine learning in concept and context drift that is relevant for understanding how participants performed in our task (Iwashita & Papa, 2018; Widmer & Kubat, 1996). For example, Devaney and Ram (1996) study small changes in category structure for the same set of stimuli. This is the same basic situation as we studied, as the overlap between categories in Figure 1 shows. They develop a COBWEB account of this sort of category drift, using a modeling framework rooted in economic models of market fluctuations. As alternative modeling approaches, Maloof (2003) develops a system that adapts by removing irrelevant examples of old concepts, Koychev (2007) presents a statistical method based on making inferences about change points, and D'Alessandro et al. (2020) develops a Bayesian model that relies on a probability distribution over possible states and the interactions between responses and feedback on each state.

Discussion

We considered a category learning experiment in which people's success in learning whether animals were healthy or diseased caused changes to those categories. Our analysis suggested that people adapt to the changing categories by preserving their knowledge about healthy animals, but discarding information about diseased animals. A model that incorporated this insight provided a better account of people's behavior than a standard incremental associative learning model.

An interesting question is why there is an asymmetry between the disease and healthy categories. It seems participants actively re-learned the diseased animals but not the healthy animals after a category change. One possible explanation for this difference is in terms of the category structures shown in Figure 1. Many animals are healthy in all of the categories, whereas only two animals are always diseased. A different, potentially complementary, explanation is in terms of the semantics of the categories. It seems natural to treat the category learning task requiring the concept of "diseased animals" to be learned. This would naturally lead to an emphasis on positive instances of the category (Tenenbaum & Griffiths, 2001), meaning animals categorized as diseased are the focus of re-learning after a category change. Further experimental work, with different category structures and category labels is needed to distinguish and evaluate these possibilities.

People learn about their world in order to guide their future decisions and actions. This means that learning can impact the world, and introduces a coupling between what people learn as their environment changes and how the environment actually does change. Thus, restricting the study of category learning, or any learning or decision-making process, to static environments or environments that change independent of people's behavior, fails to consider an important aspect of human learning.

Acknowledgments

MV was supported by UC MEXUS CONACYT doctoral fellowship 739644. SV was supported by UROP funding from the University of California Irvine.

References

D'Alessandro, M., Radev, S. T., Voss, A., & Lombardi, L. (2020). A Bayesian brain model of adaptive behavior: an application to the Wisconsin Card Sorting task. *PeerJ*, *8*, 1–32.

Dehaene, S., & Changeux, J.-P. (1991). The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex*, *1*(1), 62–79.

Devaney, M., & Ram, A. (1996). Dynamically adjusting concepts to accommodate changing contexts. In M. Kubat & G. Widmer (Eds.), *Proceedings of ICML-96 pre-conference workshop on learning in context-sensitive domains.* Bari, Italy.

El Kouari, O., Benaboud, H., & Lazaar, S. (2020). Using machine learning to deal with phishing and spam detection: An overview. In *Proceedings of the 3rd International Conference on Networking, Information Systems & Security* (pp. 1–7).

Estes, W. K. (1984). Global and local control of choice behavior by cyclically varying outcome probabilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 258–270.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633.

Gallistel, C., Mark, T. A., King, A. P., & Latham, P. (2001). The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(4), 354–372.

Iwashita, A. S., & Papa, J. P. (2018). An overview on concept drift learning. *IEEE Access*, *7*, 1532–1547.

Koychev, I. (2007). Experiments with two approaches for tracking drifting concepts. *Serdica Journal of Computing*, *1*(1), 27–44.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, *5*, 3–36.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*(2), 225–248.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*(5), 171–175.

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The cambridge handbook of computational psychology.* New York: Cambridge University Press.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1–27.

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review*, *9*(1), 43–58.

Lee, M. D., Newell, B. R., & Vandekerckhove, J. (2015). Modeling the adaptation of the termination of search in human decision making. *Decision*, *1*, 223–251.

Maloof, M. A. (2003). Incremental rule learning with partial instance memory for changing concepts. In *Proceedings of the International Joint Conference on Neural Networks* (Vol. 4, pp. 2764–2769).

Navarro, D. J., Perfors, A., & Vong, W. K. (2013). Learning time-varying categories. *Memory & Cognition*, *41*(6), 917–927.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018). A formal psychological model of classification applied to natural-science category learning. *Current Directions in Psychological Science*, *27*(2), 129–135.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*(4), 1436–1441.

Shanks, D. R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*(3), 433–443.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *13*.

Smith, D. J., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3–27.

Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General*, *139*(2), 266–298.

Sugak, A., Martynyuk, O., & Drozd, O. (2015). Models of the mutation and immunity in test behavioral evolution. In *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)* (Vol. 2, pp. 790–795).

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge (MA): The MIT Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

Westfall, H. A., & Lee, M. D. (2021). A model-based analysis of the impairment of semantic memory. *Psychonomic Bulletin & Review*, *28*, 1484–1494.

Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning*, *23*(1), 69–101.