Category Learning in Context: Modelling an Assimilation Process in Self-regulated Category Learning

Ann-Katrin Hosch (ann-katrin.hosch@uni-bremen.de)
University of Bremen, Department of Psychology, Hochschulring 18, 28359 Bremen, Germany

Janina A. Hoffmann (J.A.Hoffmann@bath.ac.uk)
University of Bath, Department of Psychology, 10W Claverton Campus, Bath, BA2 7AY, United Kingdom

Bettina von Helversen (b.helversen@uni-bremen.de)
University of Bremen, Department of Psychology, Hochschulring 18, 28359 Bremen, Germany

Abstract

Category learning, a fundamental cognitive ability, is significantly influenced by variability. In this research, we propose a model describing how people adjust information search in self-regulated category learning to the level of category variability. Participants in the self-regulated category learning task sampled from two categories until they felt confident in categorizing novel objects. Our model assumes an influence of the variability of the focal and counter category on sampling by considering a within-category and between-category processes. In both processes, variability is quantified using an information-theoretic measure. Within this model, we test if a between-category process can be better conceptualized as either a contrasting or an assimilation process. The comparison of both processes support a between-category assimilation process, where the sample size adjusts to the counter category's variability. This novel focus sheds light on between-category dynamics, providing valuable insights into the mechanisms of category learning.

Keywords: Category learning; Variability; Self-regulated learning

Introduction

Learning to categorize objects is a fundamental cognitive ability that we routinely employ in our daily lives. Is the tea you are tasting in your cup a black tea or a just a herbal infusion? Is this statement on an election poster from a left-wing or right-wing party? Is Picasso's Guernica a surrealist or cubist painting? One crucial factor shaping category learning is variability. Numerous studies in traditional category learning tasks, where participants categorize objects via feedback, reveal that high variability among encountered exemplars (i.e., category members) complicates the formation of category representations (e.g. Hahn, Bailey, and Elvin (2005); Rosch, Simpson, and Miller (1976); Homa and Vosburgh (1976)). In an ongoing research project (preregistration at https://osf.io/7u8xg) we extended this research on variability in category learning to a self-regulated task. In this task, participants actively drew category exemplars from two perceptual categories, until they felt ready to categorize novel probes to the categories in a test phase. In each trial, participants chose from which category to draw a sample, with this category designated as the "focal category" during analysis, while the other became the "counter category." We found that individuals adapted their learning based on the focal category, with higher variability leading to extended learning periods. Additionally, the counter category's variability significantly affected the learning process, with higher variability leading to more drawn samples. This paper aims to deepen our understanding of how variability in both the focal and counter categories influences category learning by proposing a computational model of the sampling process. Broadly, our model assumes that people stop sampling, when they attain a satisfactory representation of the categories. How fast this category representation can be formed depends on the category variability, with higher variability requiring more samples, modeled through an information-theoretic approach. In addition, we investigate how within-category and between-category processes can be incorporated within the model comparing two manifestations of a between-category process. Before introducing the model, we review literature emphasizing the importance of variability in evidence accumulation processes across diverse decision-making models. Then, we highlight the applications of information-theoretic measures, such as entropy, in psychological research. Finally we present our model and validate it using data from an experiment in our ongoing research project, which we will introduce in detail in the method section.

Variability in Decision-making Models

Several model frameworks propose that variability shapes how long individuals search for information when making a choice. In the models that we briefly review, variability is incorporated in various forms, including noise, uncertainty, variance, or volatility, in the evidence accumulation process. For one, the popular drift-diffusion model framework, applied to tasks ranging from orientation discrimination to categorization, treats variability as noise originating from an imperfect perception of the stimulus that influences the decision process by modulating the drift rate, i.e. the speed of evidence accumulation (Ratcliff, Smith, Brown, & McKoon, 2016). Recently, it has been shown that accounting for value uncertainty *separately* for the choice options in the drift rate, accounts best for the preferential choice task of Lee and Usher (2023).

In contrast to drift-diffusion models where accumulated evidence counts towards an option, our focus is on modeling the cessation of the information search process that precedes categorization in a self-regulated category learning paradigm. This mirrors the Decisions from Experience (DfE) paradigm, where participants sample from two options before making a choice between them. Interestingly, recent models have been

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

devised to explain the point at which individuals cease information search in the DfE paradigm, prompting the inclusion of variability in these models.

For instance, the Choice from Accumulated Samples of Experience (CHASE) model proposes that both choice and search behavior can be understood as a sequential sampling process (Markant, Pleskac, Diederich, Pachur, & Hertwig, 2015). CHASE incorporates option variability through including option variance into the drift rate, predicting longer sampling duration for higher variance options.

Similarly, the model introduced by Srivastava, Müller-Trede, Schrater, and Vul (2016), which inspired Gonzalez and Aggarwal (2021) to revise their previous, variability-unrelated stopping mechanism, emphasizes the changing focus on variability in DfE models. These models incorporate volatility, specifically in the evaluation of prediction errors, in the decision-making process regarding whether to continue sampling. Substantial changes in the prediction error prompt the individual to persist in sampling.

Having observed the pivotal role of variability in models of information search, we now turn to the role of variability in our current work.

Entropy as an Information-theoretic Measure of Variability

Departing from previous formalizations, our model employs entropy as a measure of variability, tailored to the perceptual nature of our task. Entropy is a concept from information theory and characterizes the amount of unpredictability or disorder of a system (Hirsh, Mar, & Peterson, 2012). Compared to other variability measures, entropy can account for multiple dimensions and reflects probabilistic processes. It has been integral to psychological research since the 1950s (e.g., Attneave (1954); Miller (1956)) and has found new applications in recent times, particularly in perception (Bates & Jacobs, 2020), memory (Bates, Lerch, Sims, & Jacobs, 2019), and categorization (Feldman, 2021). Entropy has also been a focus in more general models, such as the entropy model of uncertainty (Hirsh et al., 2012), which establishes a connection between entropy and psychological uncertainty. Additionally, there is research directly exploring the strong and linear relationship between the entropy of visual stimuli and subjective diversity (Stamps III, 2002). According to these theories, individuals actively work towards minimizing disorder within their immediate environment.

Modelling Self-regulated Category Learning

The self-regulated category learning task described in the introduction requires to determine at which point sufficient exemplars have been sampled to proceed to a categorization test. Our prior work suggests that both a higher category variability and a higher counter category variability increase the number of exemplars that people sample from each category. Consequently, our model assigns a pivotal role to variability in category learning. A higher variability within a category, as quantified by a lower relative entropy, determines how many samples the participants draw from each category. In addition, we aim to better understand how between-category processes shape learning. It is well-established that stimuli are not evaluated independently of their environment, showing context effects (Tourangeau & Rasinski, 1988). In the realm of category learning, comparison processes between categories can be observed when assessing how individuals adjust their learning schedules (e.g. Lu, Penney, and Kang (2021)). While established models in category learning like the prototype model concentrate solely on the focal category (Posner & Keele, 1968), and other models include a counter category without explicit emphasis (Nosofsky, 2011; Ashby & Perrin, 1988), our study explores the role of the counter category in the category learning process by explicitly modelling and comparing two between-category processes: A contrast processes and an assimilation process.

In perceptual judgment, McKenna (1984) defines contrast as an increase in the perceived difference between the contextual stimulus and the focal stimulus, while assimilation entails a decrease in this difference, pulling the judgment of the focal stimulus towards the contextual stimulus. In our task, both processes offer potential explanations on how the differing variability level of the counter category influence sampling in the focal category. Perceptually, in both processes, the differing variability level of the counter category could influence sampling in the focal category. A contrasting process would exacerbate the perception of variability in the focal category. For instance, a medium variable category would appear as fairly low variable when sampled next to a high variable counter category compared to a low variable counter category. In contrast, an assimilation process would attenuate differences leading to a medium variable category being perceived as more variable when sampled next to a high than a low variable counter category.

The variability in both, the within and between-category processes is formalized and incorporated through an entropy-related measure. Our approach, utilizing simple, systematic stimulus material and manipulation, enables the direct application of information-theoretic measures at the raw stimulus level. In the following paragraph, we outline details on the proposed model for self-regulated category learning.

Model Description

In essence, the model proposes that sampled exemplars play a role in updating the category representation, taking into account the uncertainty level associated with a category. This uncertainty governs the total number of updates needed until an individual acquires a satisfactory representation and stops sampling. The model comprises three components: (1) a function determining the decision to stop sampling based on (2) both within- and between-category processes, which are influenced by (3) variability formalized as an information-theoretic measure.

Decision to Stop Sampling

The cessation of sampling is contingent on a relative entropy term, denoted as b, encompassing the variability individuals encounter during the learning process. We integrate b into a generalized logistic function, wherein it influences temperature parameter t. If the variability is low, and thus the relative entropy term b is high, the probability to stop sampling increases more steeply than in conditions with high variability. The temperature parameter further modulates the steepness of the curve. Specifically, the probability $p(n)$ with which an individual stops sampling at a given trial is defined by the following equation:

$$p(n) = 1 - \frac{1}{(1 + e^{b \times t \times (n-0)})^{v}} \tag{1}$$

In this function, t is a temperature parameter, n is the trial number, b is the relative entropy term and v is a parameter for asymmetric curvature. The temperature parameter t captures the symmetric curvature of the function to the inflection point fixed at $z = 0$, representing the mean sample size. t thus determines the steepness of the curve and reflects when people stop sampling. The parameter v for asymmetric curvature allows the function to exhibit different curvatures on the left and right sides, mirroring observed stopping probabilities in the experiment (compare Figure 3). In our experiment, the parameter v is selected to capture a sharp rise in sampling probabilities below the mean sample size, while also representing a gradual leveling off of individuals' stopping probabilities above the mean sample size. As t captures the information when an individual ceases sampling relative to its mean sample size, t is modulated by the relative entropy term b, incorporating entropy-related measures to formalize variability in a within- and a between-category process.

Two Between-category Processes

We propose that both, a within-category process KL_A and a between-category process KL_{AB}, contribute to the general relative entropy term b that affects the decision to cease sampling. This term b has different manifestations depending on two theoretical models: the contrast model and the assimilation model.

(1) Contrast model:

The contrast model reflects the hypothesis that differences of the focal to the counter category are increased. When a higher variable category is learned with a lower variable counter category, by contrasting the variability levels, participants sample even more from the high variant category, i.e. increase their sample size. In order to model a later sampling stop, for $var(A) > var(B)$, parameter t is diminished by adding the between-category process as negative term:

$$b = KL_A + a_s \times -KL_{AB} \tag{2}$$

The parameter a_s determines the strength of the between-category process for each subject s. Conversely, the comparison of a lower variable category to a higher variable category leads to a smaller sample size, for $var(A) < var(B)$, KL_{AB} is added with a positive sign:

$$b = KL_A + a_s \times KL_{AB} \tag{3}$$

(2) Assimilation model:

In the assimilation model, differences from the focal to the counter category are decreased. Comparisons from a higher variable to a lower variable category result in a lower sample size due to assimilation to the low variability experienced in the counter category. For $var(A) > var(B)$, KL_{AB} is added with a positive sign. Conversely, the comparison of a lower variable category to a higher variable category leads to a higher sample size, thus for $var(A) < var(B)$, KL_{AB} is added with a negative sign.

Formalization of Variability in a Within- and Between-category Process

We were using an information-theoretic measure to model the uncertainty that participants encountered while learning about the categories. Within our paradigm, using a dot-distortion task, category exemplars were derived from a prototype consisting of a random dot pattern with nine dots. Variability in the category exemplars was introduced by probabilistically moving the prototype's dots in a random direction (for a detailed explanation, please refer to the "Materials" section in the Methodology). Our idea of a within- and between-category process, uses relative entropy as an information measure representing the uncertainty of a state with respect to another state. In the within-category process observing the dot in a specific location provides more certainty about the underlying category prototype compared to an uninformed uniformly distributed "baseline" prototype. For the within- and between-category processes, we utilized relative entropy or Kullback-Leibler divergence, a measure for the difference between two probability distributions. We assume that for within-category variability KL_A, individuals reduce their uncertainty when comparing the probability distribution of the focal category p_A with a distribution $p_{prototype}$ from an unknown, uniformly distributed prototype:

$$KL_A = p_A(x) * log(\frac{p_A(x)}{p_{prototype}(x)}) \tag{4}$$

While, for between-category variability KL_{AB}, individuals compare the probability distribution x of the focal category p_A with the distribution of the counter category p_B:

$$KL_{AB} = p_A(x) * log(\frac{p_A(x)}{p_B(x)}) \tag{5}$$

We evaluate our model based on data from an experiment that was part of our previous work. In the subsequent sections, we initially outline the experimental method and the modeling approach, followed by reporting behavioral analyses of the experiment and presenting the results of the modeling.

Experiment

Method

The experiment consisted of a self-regulated categorization task using a dot-distortion paradigm varying the variability of both categories.

Participants A total of 57 participants (41 female, 14 male, 2 non-binary; $M_{\text{age}} = 24.6$ years, $SD_{\text{age}} = 6.0$) voluntarily participated in the study, either for course credit or a payment of 10 Euro/hour at the University of Bremen. Prior to the commencement of the experiment, informed consent was obtained from each participant.

Procedure The experiment was implemented using jsPsych (De Leeuw, 2015) and conducted in the lab. We utilized a within-participants design, combining three variability levels for each of the two categories. This resulted in six conditions with the following combinations of category and counter-category variability levels: low-low, low-medium, medium-medium, medium-high, high-high, and low-high. The six combinations were randomized and this randomized sequence was repeated three times, resulting in participants completing a total of three blocks, each comprising six tasks of a sampling and a test phase. The prototype pair shown in a task was randomly determined. On average, participants spent approximately 38 minutes completing the experiment. During the sampling phase, participants could actively sample category exemplars from two fictitious categories, labeled 'Lum' and 'Nof', by pressing a button with the corresponding category label. Every time participants pressed a button, an exemplar from the chosen category was revealed. Participants were free to switch between categories as many times as they wished. They were instructed to continue sampling from categories 'Lum' and 'Nof' until they felt confident in categorizing the exemplars into the fictitious categories. Once they reached that point, they ended the sampling phase with a button press and proceeded to the subsequent test phase. In the test phase, participants indicated the likelihood that the exemplar belonged to either category for eight novel stimuli of different variability levels. The fictitious category labels as well as the prototypes were randomly assigned to the left or right side of the sampling screen display for each participant.

Material Random dot patterns served as stimuli. Prototypes for categories A and B were generated from one joint prototype. This joint prototype was generated initially by placing nine dots randomly with a minimum distance from one another. Subsequently, two versions were created by moving every dot with a defined distance in a random direction (one from nine possible directions) for prototype A and in the opposite direction for prototype B. To implement low, medium, and high variability within a category, we created exemplars for a category by distorting the prototype, with each of the nine dots moving with probabilities of $p_{dot} = .1$, $p_{dot} = .2$, and $p_{dot} = .3$ respectively. Figure 1 illustrates the variability manipulation. Sixty exemplars were generated from each prototype, and during the learning phase of the experiment, exemplars were randomly drawn from this pool.

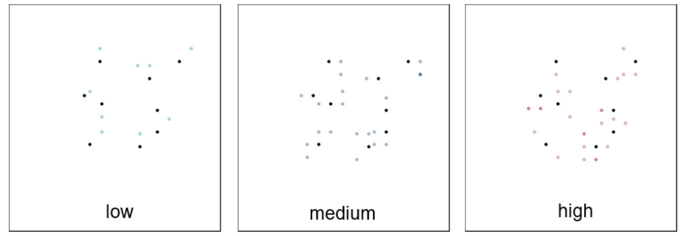

Figure 1: Illustration of the variability manipulation

Note. Shifts from the original black dots (i.e. the prototype) in the low (green), medium (blue) and high (red) conditions. The colored dots simulate possible locations of the dot when moving with a certain probability, i.e. according to the variability level, for a number of overlapping plotted stimuli.

Each dot can appear in nine possible locations that differ in the probability that the dot appears in them. The dot has the highest probability of appearing at the original location of the dot (i.e. the location in the prototype), while the other locations are associated with the condition-dependent probability that the dot moves to this location (each divided by the number of possible locations). Specifically, the distribution takes the following form for a variability condition with $p_{dot} = .1$, $p_{dot} = .2$, and $p_{dot} = .3$: $p = [1 - p_{dot}, \frac{p_{dot}}{8}, \frac{p_{dot}}{8}, \frac{p_{dot}}{8}, \frac{p_{dot}}{8}, \frac{p_{dot}}{8}, \frac{p_{dot}}{8}, \frac{p_{dot}}{8}, \frac{p_{dot}}{8}]$. Relative entropy compares the variable category exemplars to an uniformed "baseline" prototype, resulting in a high relative entropy value when the probability of dots moving is low, and a low relative entropy value when the probability is high (see the Model Description section).

Model Fitting

To gain a detailed understanding of when people decide to stop sampling, we applied the model described in the preceding paragraph. In the model, we fitted parameters hierarchically, for the asymmetric curvature and the temperature across conditions (i.e. on group-level) and the parameter a_s which determines the strength of the between-category process, for each individual. As priors for the asymmetric curvature and the temperature parameter, we chose $v \sim N(0, \frac{1}{5^2})[0,]$ and $t \sim N(0.25, \frac{1}{0.1^2})[0,3]$. Further, we estimated the parameter a_s for each individual with the prior $a_s \sim N(a, \frac{1}{5^2})[0,]$ and at the group-level with the prior $a \sim N(0, \frac{1}{5^2})[0,]$. We fitted the model with JAGS (Plummer et al., 2003) and the package *r2jags* (Su & Yajima, 2021) in R (Version 4.1.2) (R Core Team, 2021). We ran each model with two chains and 10,000 iterations. All models converged with $\hat{R} < 1.01$ for all parameters and a lowest effective sample size of $ESS = 2200$.

In addition, we aimed to compare the models with plausible alternatives. One such alternative involved fitting a model without a between-category process. Furthermore, we implemented all models by formalizing variability solely as raw probability, specifically the probability of a dot staying in its location, rather than using the information-theoretic measure. To facilitate model comparison and identify the best-fitting model, we used posterior model probabilities. The model selection variable was drawn from a categorical distribution, assuming equal probabilities for each model.

Results

Behavioral Analysis

Sample size per category was on average $M = 14.22$ ($Md = 11$, $SD = 10.75$). Due to high inter-individual differences in the average sample sizes, we report and use standardized sample sizes for the analyses. We calculated standardized z-scores for the sample sizes by subtracting the participant's mean sample size from the individual sample sizes and divided by the standard deviation in each of the conditions. Figure 2 illustrates the standardized sample sizes for all variability levels. Notably, there was a discernible increase in sample sizes with variability levels, progressing from low ($z = -0.23$, $SD = 0.83$) to medium ($z = -0.03$, $SD = 1.02$) to high ($z = 0.25$, $SD = 1.11$) for conditions with equal variability levels.

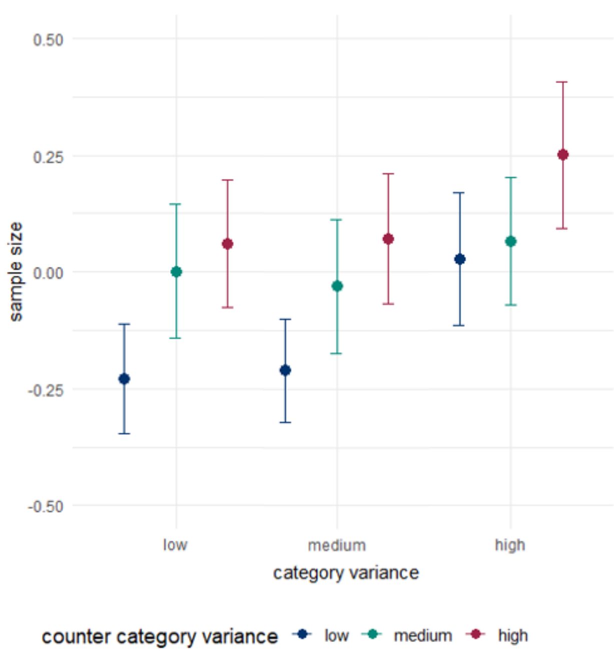

Figure 2: Sample size

Note. Plotted are the mean standardized sample sizes for the focal and counter category's variability levels. Error bars denote the standard deviation.

To analyze how category variability affects the length of information search as measures by sample size, we calculated a linear mixed model with the following formula:

sample size $\sim$ category variability * counter category variability + block + (category variability * counter category variability || subject) + (1 | prototype pair)

We identified a significant main effect of category variability, $b = 1.19$, $F(1, 54.30) = 7.37$, $p = .009$. Participants tended to draw more samples as the variability of the category increased. Additionally, a significant effect was observed for the counter-category variability, $b = 1.83$, $F(1, 1620.75) = 18.65$, $p < .001$, indicating that participants also drew more samples when the non-focal counter category exhibited higher variability. However, the interaction of variability levels did not reach significance ($b = 0.92$, $F(1, 48.89) = 0.47$, $p = .498$). Thus, inferential statistic results suggest that both, category and counter category have an impact on sample size.

Modelling Results

The model selection in which we had included the two suggested models with a contrast and an assimilation process, a model without between-category process and all models with raw probabilities instead of information-theoretic formalization of variability, suggested that the assimilation model with the information-theoretic formalization described the data best in all observations. Specifically, for conditions with differing variability levels, the selection variable favored the assimilation model with the information-theoretic formalization. A comparison of the information-theoretic model with *DIC* values further supported the results: $DIC_{assimilation} = 36316.1$, $DIC_{contrast} = 37544.7$ and for the model without between-category process, $DIC = 37451.7$.

In Figure 3 we illustrate the model fit for two conditions. The orange line depicts the model without the assumption of a between-category process, but only the variability that is experienced within the category. From the blue and purple line, we can infer how the between-category process influences stopping probability: When a category with a low variability is learned with a highly variable counter category, the contrast process predicts earlier and the assimilation process predicts later stopping than the model without between-category process. Vice versa, when a high variable category is learned with a low variable counter category, the contrast process assumes later stopping while the assimilation model assumes earlier stopping, again compared to a model without between-category process. We can see that the data is better described if we assume that a assimilation process takes place. Specifically, for conditions in which the counter category has a higher variability level than the focal category (left graph), the data is better described by the assimilation model that assumes later stopping, while for the graph on the right hand side in which the counter category has a lower variability than the focal category, the data is better described by the assimilation model that assumes earlier stopping.

The fitted parameter a_s which describes the strength of the between-category process, varied considerably between indi-

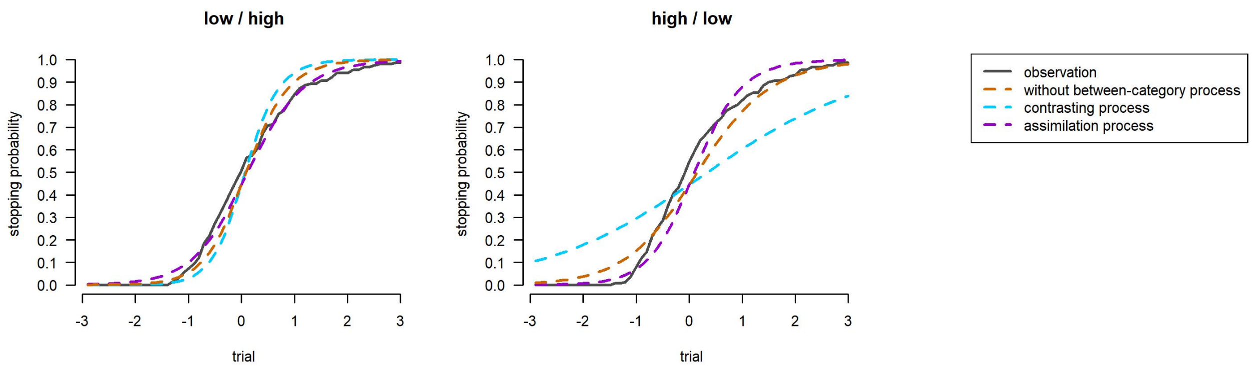

Figure 3: Model fit

Note. Plotted are the fitted (broken line) and observed (solid line) probabilities to stop sampling for the z-transformed trial numbers for two of the conditions (low focal category variability/high counter category variability and vice versa). The graph illustrates fitted values for both the contrast model (blue) and the assimilation model (purple), along with the model without a between-category process for comparison (orange). Parameter a was set to $a = 3.52$ in both models, to illustrate the influence of the between-category process.

viduals, $M = 3.52$ and $SD = 2.40$.

Discussion

In this study, we modeled sampling behavior in a novel self-regulated category learning task. We employed an information-theoretic measure to represent variability, underscoring the relevance of information theory in perceptual (learning) processes. Our conceptualization posits that category learning involves both within-category and between-category processes, both driven by variability and jointly influencing the decision to terminate sampling. For the between-category process, we explored two potential mechanisms: a contrasting process, where category characteristics are juxtaposed during learning, and an assimilation process, where learning adapts to the counter category's characteristics. Our model fitting results revealed that category learning is not solely influenced by a within-category process; the between-category process also plays a crucial role. Participants adjusted their sampling behavior based on the counter category's variability, indicating an assimilation between-category process. The model fitting results are substantiated by descriptive observations, providing additional insights into the significant impact of both the category and counter category found in our generalized linear mixed model analysis. This assimilation process is clearly illustrated in the observed pattern presented in Figure 2. Specifically, sample sizes of a low-variable category increased in the presence of a high-variable counter category, while sample sizes of a high-variable category decreased in the presence of a low-variable counter category.

Nevertheless, our model has certain limitations. While fitting the variability term on the temperature parameter of a generalized logistic function captures sampling stop behavior relative to the mean sample size, it does not account for other central characteristics of sampling behavior, such as a shifting mean sample size depending on the condition. The next version of the model will address this shortcoming. Additionally, future research should explore the generalizability and applicability of an information-theoretic measure in other category learning tasks that employ different materials.

Despite these limitations, the identification of an assimilation process holds significant implications for understanding the mechanisms of category learning. It suggests that individuals perceive both categories as part of a unified entity to be learned, overshadowing contrasting processes that typically emphasize differences. While previous evidence supports the idea that comparison processes can enhance category learning (Kurtz, Boukrina, & Gentner, 2013), and individuals actively compare categories, particularly when categories are similar (Lu et al., 2021), there has been limited emphasis on between-category processes in category learning research. This study aims to address this gap and shed light on how the counter category is incorporated into the learning process. The observed assimilation towards the counter category's variability not only suggests mutual influence between their characteristics but also implies a close interconnection between both categories. This connection suggests that learning may extend beyond the focal category, encompassing insights about the broader category context.

References

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*(1), 124-150.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183-193.

Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological Review*, *127*(5), 891-917.

Bates, C. J., Lerch, R. A., Sims, C. R., & Jacobs, R. A. (2019). Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *Journal of Vision*, *19*(2), 11.

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.

Feldman, J. (2021). Mutual information and categorical perception. *Psychological Science*, *32*(8), 1298–1310.

Gonzalez, P., Cleotilde, & Aggarwal, P. (2021). *Sequential decisions from sampling: Inductive generation of stopping decisions using instance-based learning theory.* PsyArXiv. doi: 10.31234/osf.io/t4vmb

Hahn, U., Bailey, T. M., & Elvin, L. B. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*(2), 289–302.

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304-320.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(3), 322-330.

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1303-1310.

Lee, D. G., & Usher, M. (2023). Value certainty in drift-diffusion models of preferential choice. *Psychological Review*, *130*(3), 790–806.

Lu, X., Penney, T. B., & Kang, S. H. (2021). Category similarity affects study choices in self-regulated learning. *Memory & Cognition*, *49*, 67–82.

Markant, D., Pleskac, T. J., Diederich, A., Pachur, T., & Hertwig, R. (2015). Modeling choice and search in decisions from experience: A sequential sampling approach. In *Proceedings 37th annual meeting of the cognitive science society* (pp. 1512–1517).

McKenna, F. P. (1984). Assimilation and contrast in perceptual judgments. *The Quarterly Journal of Experimental Psychology Section A*, *36*(3), 531–548.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In M. Pothos & A. Wills (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge University Press.

Plummer, M., et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, pp. 1–10).

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3p1), 353-363.

R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491-502.

Srivastava, N., Müller-Trede, J., Schrater, P. R., & Vul, E. (2016). Modeling sampling duration in decisions from experience. In *Proceedings of the 38th annual meeting of the cognitive science society.*

Stamps III, A. E. (2002). Entropy, visual diversity, and preference. *The Journal of General Psychology*, *129*(3), 300-320.

Su, Y.-S., & Yajima, M. (2021). R2jags: Using R to run 'JAGS' [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=R2jags` (R package version 0.7-1)

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, *103*(3), 299-314.