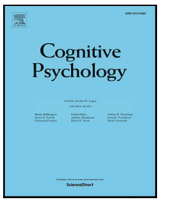
Cognitive Psychology 152 (2024) 101670

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych



Task imprinting: Another mechanism of representational change?

Mirko Thalmann [a,*], Theo A.J. Schäfer [b], Stephanie Theves [b], Christian F. Doeller [b], Eric Schulz [a]

[a] Max Planck Institute for Biological Cybernetics, Max–Planck-Ring 8, 72076 Tübingen, Germany
[b] Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1, 04303 Leipzig, Germany

ARTICLE INFO

Dataset link: https://osf.io/pgyr4/files/, https://osf.io/pgyr4/registrations/

Keywords:
Representations
Representational change
Category learning
Memory

ABSTRACT

Research from several areas suggests that mental representations adapt to the specific tasks we carry out in our environment. In this study, we propose a mechanism of adaptive representational change, *task imprinting*. Thereby, we introduce a computational model, which portrays task imprinting as an adaptation to specific task goals via selective storage of helpful representations in long-term memory. We test the main qualitative prediction of the model in four behavioral experiments using healthy young adults as participants. In each experiment, we assess participants' baseline representations in the beginning of the experiment, then expose participants to one of two tasks intended to shape representations differently according to our model, and finally assess any potential change in representations. Crucially, the tasks used to measure representations differ in the amount that strategic, judgmental processes play a role. The results of Experiments 1 and 2 allow us to exclude the option that representations used in more perceptual tasks become biased categorically. The results of Experiment 4 make it likely that people strategically decide given the specific task context whether they use categorical information or not. One signature of representational change was however observed: category learning practice increased the perceptual sensitivity over and above mere exposure to the same stimuli.

1. Introduction

How does our cognitive system react when we learn to perform a task, for example to categorize a flavor as belonging to a fruit. A simple solution is to rely on a strategy that we already know. For example, an allergic reaction against the specific fruit may tell us that it is a Kiwi. A harder solution is to learn a strategy, which maps fruit taste profiles to fruit identities, from the ground. This could be achieved, for example, by using the rule that a highly acidic fruit is a grapefruit (Ashby & Gott, 1988), by using individually stored episodes of fruits previously eaten to generalize to the current fruit (Nosofsky, 1986), by using representations of prototypical fruits and go with the closest prototype (Homa et al., 1982; Minda & Smith, 2001), or by using a combination of the latter two (Vanpaemel & Navarro, 2007). The speed of the involved basic cognitive processes may even increase when we use the same strategy over extended periods (Case et al., 1982). A different, intriguing idea is that the representations of the individual fruits themselves change with learning.

The idea of representation learning has gained a lot of traction in the field of machine learning in the last decades. The reason for that is that the performance of machine learning algorithms in terms of predictive accuracy heavily depends on the representation of the data (i.e., the features Bengio et al., 2014). For example, the success of speech recognition models depended to a large degree

* Correspondence to: Research Group Computational Principles of Intelligence, Max Planck Institute for Biological Cybernetics, Max–Planck-Ring 8, 72076 Tübingen, Baden-Württemberg, Germany.
E-mail address: mirkothalmann@hotmail.com (M. Thalmann).

https://doi.org/10.1016/j.cogpsych.2024.101670
Received 16 August 2023; Received in revised form 19 April 2024; Accepted 25 June 2024
Available online 13 July 2024

on learning good representations of speech signals. By analogy, it could be assumed that achieving high performance on a given task by humans also depends on learning good mental representations. A main question to be tackled then is how representations become more favorable to carry out a task. Or in other words, how do representations change? Research targeted to understanding mental representations and their change broadly comes from three fields: neuroscience, cognitive science and cognitive psychology. Despite continued research in these areas and findings interpreted in favor of the idea that representations of individual objects change (Dubova & Goldstone, 2021; Goldstone et al., 2001; Karagoz et al., 2022) it is currently not clear what – if anything – changes with representational change.

We start this article with a literature review summarizing different lines of previous research targeting the topic of representational change. This allows us to situate a new suggested mechanism of representational change, referred to as *task imprinting*, within the landscape of existing empirical and theoretical work. We then implement the mechanism as a computational model, which portrays representational change as a response to task-specific practice. We further develop a behavioral paradigm to test whether the proposed mechanism of representational change can be measured behaviorally. We eventually test the qualitative predictions of the model in a series of four controlled online experiments, three of them pre-registered.

The results are mostly in disagreement with the model's predictions. They show that representations of individual objects do not become biased as expected according to practice in a category learning task. Practice in a category learning task, however, tended to increase the precision of the individual object representations, over and above the increase in precision in a control task. Over the progression of the four experiments we varied the task intended to measure representations. When the task was designed to measure predominantly perceptual details about objects, participants' responses did not change as predicted by the task imprinting model. However, when the task highlighted higher-level information about the stimulus, such as category membership, participants' responses changed accordingly. The results are consistent with the idea that participants strategically incorporate learned categorical representations in addition to individual object representations when a task context is sufficiently similar to the category learning task.

2. Background

The majority of scientific results about representational change comes from three areas. In the area of perceptual learning, researchers are interested in uncovering the factors that determine how entire stimuli or individual feature dimensions are mentally represented. A major question in the area of category learning and concept learning is when people use abstracted representations, such as rules or prototypes, and when they use individually stored instances to represent categories and concepts. Furthermore, research about reinforcement learning in humans is interested in understanding how repeated exposure to reward signals shapes the construction of cognitive maps. To locate our proposed mechanism of task imprinting, we are summarizing the main conclusions of these areas aligned with empirical evidence in the following.

Goldstone (1998) discriminates between four mechanisms of perceptual learning: attentional weighting, stimulus imprinting, unitization, and differentiation. *Attentional weighting* refers to the fact that people learn to attend to stimulus dimensions, which are relevant for a given task at hand, with a higher weight. For example, the generalized context model for categorization tasks (Nosofsky, 1986) assumes that selective attention can be distributed across the feature dimensions of a stimulus to be classified. When people learn to pay more attention to a particular dimension, they zoom into that dimension and represent it in a more fine-grained way. Kruschke (2005) showed that selective attention is a necessary ingredient for connectionist cognitive models to be able to mimic learning difficulties in three classical learning paradigms. A similar explanation of attentional weighting is offered by rate–distortion accounts of perception (e.g., Bates & Jacobs, 2020; Sims, 2016). They posit that limited cognitive resources are optimally distributed to minimize the cost of a categorization error. Therefore categorization relevant dimensions should be represented more precisely. In models, which assume trial-by-trial variability in the perceptual information obtained from every object (e.g., Ashby & Lee, 1993), attention towards a dimension reduces the perceptual variability on that dimension (see for example Braida & Durlach, 1972; Durlach & Braida, 1969; Luce et al., 1976, 1982). There is also evidence from neuroscience that hippocampal object representations during categorization are modulated by attention to the currently decision-relevant feature (Mack et al., 2016, 2018). Furthermore, Theves and colleagues show that the hippocampus specifically integrates categorization relevant feature dimensions in a common representational space, thereby mapping distances between exemplars and category boundaries (Theves et al., 2019, 2020).

Attentional weighting is tightly linked to the phenomenon of categorical perception. It describes the observation that two stimuli separated by a fixed distance in feature space are harder to discriminate when they come from the same category than when they come from different categories. It was first observed in the domain of speech perception. Liberman et al. (1957) found that discrimination between auditorily presented phonemes is more difficult when they come from the same phoneme category (e.g.two b's) than when they come from different phoneme categories (e.g. b vs. d). While the emergence of categorical perception for natural speech might be targeted from a developmental perspective, several studies examined it experimentally in the visual domain. Goldstone (1994) showed that perceptual sensitivity to discriminate between visually presented squares varying in size and brightness increases after category learning for stimuli assigned to different categories. The sensitivity increased for the category-relevant dimension but did not decrease for the irrelevant one, which is halfway consistent with the idea of categorical perception. Goldstone et al. (2001) showed that within-category items were not rated as more similar after category learning than before, but surprisingly between-category items were. Consistent with their categorical perception account was the observation, though, that the difference of similarity ratings from two stimuli to a neutral stimulus became smaller, when the two stimuli came from the same category, but not, when they came from different categories.

Theoretically, categorical perception can be explained by disproportionate attention and perceptual sensitivity close to the category boundaries (Goldstone, 1998) or an attractor state in a hidden layer of a connectionist model (Harnad, 1995), which deemphasizes differences between representations of stimuli from the same category. Categorical perception can also be explained within a Bayesian framework. In that case, categorization responses are considered to be a mixture of the prior probability of the category and the likelihood of the stimulus. Bayesian models typically predict an attraction of representations towards category prototypes and therefore decrease distances between representations within the same category, but increase distances between representations between categories (e.g., in the Category Adjustment Model Huttenlocher et al., 2000).

Evidence for the prediction of attraction to prototypes has been shown in several studies. For example, Huttenlocher et al. (1991) presented participants with a location on a circle to remember for immediate recall. An important finding was that participants' responses were biased towards certain stereotypical angles (45 degrees, 135 degrees and so on) and towards locations halfway between the center of the circle and its circumference. In a similar vein, Hasantash and Afraz (2020) presented participants with one target color patch in one of two experimental conditions. In the simultaneous matching condition, participants adjusted a variable color patch to match the simultaneously presented target color patch. In the sequential matching condition, they adjusted the variable color patch after the target patch had disappeared for a ten second retention interval. At the end of the experiment participants were given a color naming task. The authors found that the precision of participants' responses in the sequential matching task but not in the simultaneous matching task correlated with the size of the color vocabulary. Moreover, precision was higher in regions of the color space where the average density of the color vocabulary was higher.

Stimulus imprinting refers to the process by which repeated exposures (and possibly responses) to the same stimulus lead to more accurate and faster perception of that stimulus. On the theoretical level, it can be explained by referring to memory storage of exemplars (Nosofsky, 1986) or instances (Logan, 1989) or by the adaptation of sensory processes to the repeated exposure to the same stimulus (Goldstone, 1998). For example, participants categorize stimuli, which have been presented to them more often, with higher accuracy (Nosofsky, 1991). According to Shiffrin and Schneider (1977), at least a part of that effect may be due to a reduced demand of controlled attention, which is assumed to be capacity limited. In their studies, they showed that less controlled attention is required to respond to a consistent mapping of a stimulus set to a response than to a varied mapping. In that series of experiments, more automatic processing, which is assumed to be the flip side of controlled attention, also led to more accurate and faster responses.

Similarly, *unitization* describes the process by which stimuli initially perceived separately become perceived together as one unit. For example, Blaha et al. (2009) asked participants to categorize complex stimuli consisting of five different "squiggly" line segments, which were sampled from a pool of 16 line segments. Participants initially processed these line segments sequentially, but learned to process them in parallel after extensive training. That is, the authors showed that a measure of workload capacity, reflecting the amount of information integrated in a fixed time window, substantially increased after a seven-day training. Moreover, chess experts perceive mid-game positions as familiar sub-configurations of pieces (i.e., chunks), whereas novices perceive these positions on the level of individual pieces (Chase & Simon, 1973). Thalmann et al. (2019) showed that well learned chunks require less working-memory capacity and therefore allow people to focus on more relevant information at the same time. A related study about the construction of cognitive maps in humans (Karagoz et al., 2022) showed that people perceive stimuli as more similar when they share an associated reward, but not when they are merely presented simultaneously. The latter study suggests that a shared reward may render the representations of two stimuli more similar, possibly to effectively collect rewards in an environment.

Finally, *differentiation* describes the observation that stimuli or feature dimensions, which are initially perceived as the same, become represented separately over time. For example, training to discriminate between different types of wine led to a 14% improvement in a same-different task (Owen & Machamer, 1979). Similarly, Lively et al. (1993) showed that Japanese speakers can be trained to differentiate between the phonemes /r/ and /l/, which are not present in Japanese. Goldstone and Steyvers (2001) showed that participants learned to differentiate integral dimensions given training conditions that emphasize the orthogonality of the dimensions.

The four mechanisms of perceptual learning suggest that perception and categorization are closely related. While the previously presented studies focused on perceptual learning of individual stimuli, there is also evidence for representational change on the level of the representations of categories. Theories of category learning can be roughly characterized by the extent that they rely on representations of individually presented stimuli (e.g., Nosofsky, 1986), on representations of rules (e.g., Ashby & Gott, 1988), or on representations of category prototypes (e.g., Posner & Keele, 1968). Several authors suggested that there are shifts in the level of abstraction of representations people use to categorize objects in their environment. For example, Smith and Minda (1998) argued that there is a shift from an initial prototype-based process to a later process relying on individually stored exemplars when participants are exposed to an extended period of category learning. Johansen and Palmeri (2002) argued that the shift may be described better as a shift from a hypothesis-driven rule-based categorization process towards the exemplar process.

Even though we framed this initial literature review around mechanisms of perceptual learning, we have to mention here that locating the cause of changed responding at the perceptual level is hardly ever possible. For example, increased precision of perceptual representations on a categorization-relevant dimension can be explained in at least two ways. First, the effect is consistent with an account of changed perceptual representations (i.e., increased precision of perceptual representations). Second, the effect is also consistent with an account of dimensional attention with the perceptual representations remaining unchanged. Similar considerations apply to the other presented mechanisms. Here, we want to acknowledge that making clear-cut statements about the location of representational change are not possible with the current experimental and modeling setups. We will come back to this issue in the modeling section.

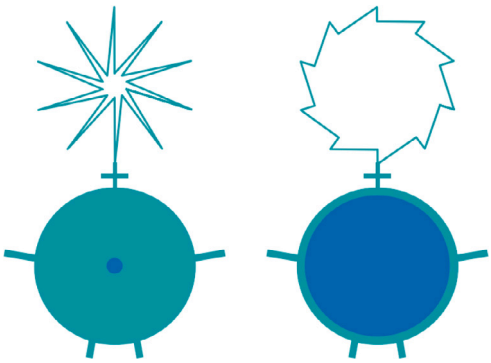


Fig. 1. Left: Stimulus with minimum values on both feature dimensions. Right: Stimulus with maximum values on both feature dimensions.

To summarize, research about perceptual learning, category learning, and human reinforcement learning can be considered as evidence that there are representational shifts of whole stimuli, parts of feature dimensions or entire feature dimensions, categories, and the perceived similarity between stimuli. There is also evidence of systematic representational biases when people reconstruct information from memory as compared to perceiving it. One possibility is that the underlying force of representational change is adaptivity to the environment (Anderson, 1991; Simon, 1996). For example, focusing on a categorization-relevant dimension is favorable, because it improves performance. In the current paper, we add to the growing body of research about representational change by proposing an additional mechanism of representational change: *task imprinting*. The basic idea is that our cognitive system preferentially stores representations in memory, which are favorable for a given task. In the following, we introduce a controlled setup, which allows us to observe such a representational change in an experimental behavioral setting. We then introduce a computational model, which makes specific predictions for that setup. We then explain how we test the model predictions in a series of four experiments.

3. A model of task imprinting

In the following, we introduce a computational model that portrays task imprinting as an adaptive response to task-specific practice. In order to explain the model and simulation results of the model we first give a brief sketch of the experimental setup. Knowledge about the latter should help to understand the decisions we took in our modeling approach.

3.1. Experimental setup

An important point to show in our approach is that representations change differently according to different tasks. We chose two tasks that differed in their respective goals. One task was a category learning task, in which biased representations are particularly helpful. With biased representations we refer to representations that systematically deviate in their mean as opposed to the ground truth. For example, a helpful bias in the category learning task would be to represent any stimulus with the category prototype. This would massively improve categorization accuracy, given any type of categorization model. The other task was a sequential comparison task. Biased representations are not helpful in the latter, because representing certain pairs of objects with means closer to each other than other pairs does not improve performance in the sequential comparison task. In both tasks, participants were confronted with the same 100 two-dimensional stimuli. The two dimensions of the stimuli were represented as the spikiness of the head and the fill of the belly of "monsters" (see Fig. 1). We defined the 100 stimuli according to equally-spaced locations in the two-dimensional feature space (see Fig. 2). In the category learning task, participants were required to learn to assign these stimuli into two categories (Experiment 1, category structure similar as in Milton & Pothos, 2011; Nosofsky et al., 2005) or into four categories (Experiments 2–4, similar as in Nosofsky et al., 2005). In the sequential comparison task, participants were required to judge how similar two consecutively presented stimuli were to one another.

All experiments followed the same sequential logic. Before and after practice in one of those tasks we measured representations. Measuring representations before and after task practice with the same stimuli but different goals allowed us to get an estimate of representational change by the logic of subtraction. Any task-specific differences can be attributed to representational change due to the different goals.

The way we measured representations varied across experiments, though. In Experiments 1 and 2, we used a continuous reproduction task (Pertzov et al., 2013; Souza et al., 2014). In Experiment 3, we used a simultaneous comparison task, in which participants were required to rate how similar two stimuli were to each other. In Experiment 4, we used a simultaneous comparison task, in which participants were required to rate how likely two stimuli were to come from the same category. We chose these different tasks, because they differ in the amount that strategic, judgmental processes play a role (Firestone & Scholl, 2016). Whereas the influence of these processes in the continuous reproduction task is present, but small (Goldstone, 1994), the influence tends to be non-negligible when participants are asked to say how similar two simultaneously presented stimuli are they have already encountered during category learning (Experiment 3 Goldstone et al., 2001). Finally, the judgmental influence is largest for the task used in Experiment 4, in which participants were asked to rate how likely two stimuli are to come from the same category.

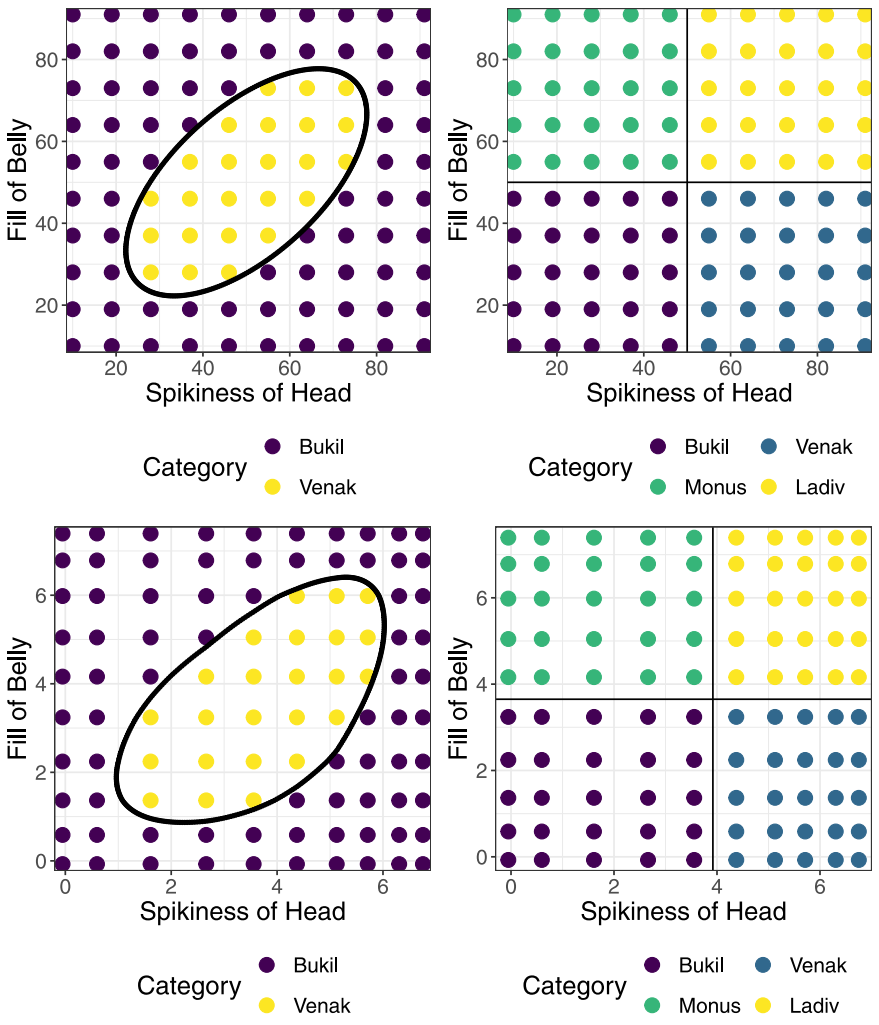


Fig. 2. Two-dimensional category structures in Experiment 1 (left) and Experiments 2–4 (right). Top row refers to object space, bottom row refers to psychological space.

3.2. Model specification

We make the assumption that there is a difference between a perceptual representation and a memory-based representation. The perceptual representation and its associated variability are assumed to be driven by the stimulus in the environment and the signal transmission to early sensory areas (e.g., Ashby & Lee, 1993; Ashby & Townsend, 1986; Bays, 2014; Pouget et al., 2000). The memory-based stimulus representation is similar to memory for individually stored exemplars or instances (Logan, 1989; Nosofsky, 1986), but we assume there is some task-based gate, which decides whether an exemplar or instance is finally stored in memory. The computational modeling is targeted to the tasks, which are assumed to shape memory representations. We therefore focus on modeling the representational change caused by the category learning task and the sequential comparison task. The result of that process is that some representations are preferentially stored in long-term memory. And a key assumption of the model is that these preferentially stored long-term memory representations are also behaviorally relevant in the task aimed at measuring representations (i.e., after the category learning task or the sequential comparison task; e.g., Souza et al., 2021). The computational model consists of three sequential stages we refer to as the perceptual representation, the task mapping, and the long-term memory gate.

3.2.1. Perceptual representation

In the first stage, one stimulus out of the set of 100 stimuli is presented. The objective means in the artificial feature space were created by crossing 10 equally-spaced values in each dimension:

$$\boldsymbol{\mu}_i = [i, j], \quad with \quad i = 1:10 \quad and \quad j = 1:10 \tag{1}$$

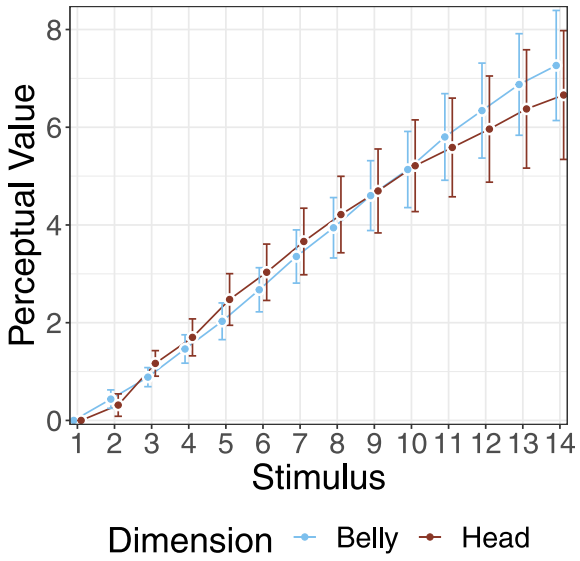


Fig. 3. Average psychological representations for the 14 linearly spaced objective stimulus values derived from the psychophysical experiment. Note. Error bars represent 95% confidence intervals.

We derived the subjective, psychological means of the perceptual distribution in a psychophysical scaling experiment, the methods and results of which are reported in the Appendix. On average, the two dimensions were resolved with an about equal perceptual precision on both dimensions (see Fig. 3).

We then modeled the trial-wise perceptual representation via a sample from a multivariate normal with uncorrelated variances

$$\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} = \mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \sigma^2 \mathbf{I}_2), \quad i = 0, \ldots, 99 \tag{2}$$

The univariate means μ_i of the 10 shown stimuli in the subjective feature space were inferred by linearly interpolating the respective mean values from the psychophysical experiment. The 10 univariate means on each dimension were then fully crossed to obtain the subjective representations of the 100 stimuli shown in the bottom panel of Fig. 2. The variance σ refers to variability of the perceptual distribution $sd_{perceptual}$. It has to be noted, that we make the explicit assumption that the perception of the objects is probabilistic. This assumption is in the tradition of population coding models (e.g., Bays, 2014; Pouget et al., 2000) and consistent with previous theories of perceptual variability (Ashby & Townsend, 1986). However, whether perception eventually is probabilistic or not is an unresolved question (e.g., Rahnev et al., 2021; Rahnev & Denison, 2018).

3.2.2. Task mapping

In the second stage, perceptual representations are mapped to responses in the category learning task and in the sequential comparison task. For the category learning task, we used three different models of category learning: an exemplar model, a prototype model, and a rule-based model. We used the generalized context model as a representative exemplar model (Nosofsky, 1986), a naïvely Bayesian classifier (e.g., John & Langley, 2013) for Experiment 1, the multiplicative prototype model from Nosofsky and Zaki (2002) for Experiments 2–4 as prototype models, and a rule-based categorization model inspired by the general recognition theory (Ashby & Townsend, 1986) with the implementation similar as in the study by Maddox et al. (2004). Each model takes the feature values x_i of the perceptual representation of stimulus on trial i on the two dimensions as an input and returns category probabilities $p(cat_k|\ x_i)$ for every category k out of m categories.

3.2.2.1. Exemplar model. The exemplar model arrives at those probabilities by dividing the summed similarity from the current stimulus i to all stimuli categorized into the same category by the summed similarity from the current stimulus to all observed stimuli.

The similarity between the representations of two stimuli is an exponentially decaying function of the distance d between the two stimuli

$$\eta_{ij} = \exp^{-d_{ij}}. \tag{3}$$

Let n_k be the number of stimuli that fall under the category k and m be the number of categories, then

$$p_e(c_k|\boldsymbol{x}_i) = \frac{\sum_{j=1}^{n_k} \exp(-d_{ij})}{\sum_{l=1}^{m} \sum_{j=1}^{n_l} \exp(-d_{ij})}, \tag{4}$$

where $d_{ij} = c \cdot \sqrt{w_1(x_{i1} - x_{j1})^2 + (1 - w_1)(x_{i2} - x_{j2})^2}$ s.t. $w_1 = [0; 1]$.

The individual w_d values represent the attention weight put on each of the two feature dimensions. c is a sensitivity/generalization parameter, which controls the slope of the negatively decaying exponential function. Because the same number of stimuli was presented in all categories, and for simplicity, we omitted a response bias parameter.

3.2.2.2. Prototype models. The *naïve Bayes* prototype model arrives at the probability that stimulus x_i is categorized into category k by calculating

$$p_{pn}(c_k|\boldsymbol{x}_i) = \frac{p(\mathbf{x}_i|c_k)p(c_k)}{\sum_{l=1}^{m} p(\mathbf{x}_i|c_l)p(c_l)}, \tag{5}$$

where the likelihood $p(\mathbf{x}_i|c_k) = \mathcal{N}\left(\begin{bmatrix} m_{k1} \\ m_{k2} \end{bmatrix}, \begin{bmatrix} \sigma_{k1}^2 & 0 \\ 0 & \sigma_{k2}^2 \end{bmatrix}\right)$.

Thereby, $p(c_k)$ represents the prior probability of a stimulus to belong to category k. Similarly as above, we omitted any category response bias, and set this parameter to the proportion of stimuli from category k in the training set. Every category is represented by independent normal distributions on both feature dimensions; their means and variances are fit as parameters to maximize the likelihood of the stimuli in each category. The conditional probability $p(x_i|\ c_k)$ is then calculated by multiplying the probability densities of two independent normal distributions:

$$p(\boldsymbol{x}_i|\ c_k) = \prod_{d=1}^{2} \mathcal{N}(\mu_d, \sigma_d) \tag{6}$$

As the exemplar model, the *multiplicative prototype* model operates on the level of similarity representations. It uses a mixture between random guessing and the similarity of a presented stimulus with the stored prototypes of all categories to select one of the response options:

$$p_{pm}(c_k|\boldsymbol{x}_i) = \frac{g}{m} + (1 - g)\frac{\exp(-cd_{ik})}{\sum_{l=1}^{m} \exp(-cd_{il})}, \tag{7}$$

where $d_{ik} = c \cdot \sqrt{w_1(x_{i1} - P_{k1})^2 + (1 - w_1)(x_{i2} - P_{k2})^2}$ s.t. $w_1 = [0; 1]$ and $g = [0; 1]$.

Note that P_{k1} reflects the feature value on dimension 1 for the prototype of category k.

3.2.2.3. Rule-based model. We applied the rule-based model only for the square category structure in Experiments 2–4. It arrives at the probability that a presented stimulus x_i belongs to category j by calculating

$$p_r(c_k|\boldsymbol{x}_i) = \int_{\boldsymbol{a}(\boldsymbol{k})}^{\boldsymbol{b}(\boldsymbol{k})} \mathcal{N}\left(\mathbf{x}_i, \sigma^2\mathbf{I}_2\right) \mathrm{d}\mathbf{x}_i, \tag{8}$$

where the vectors $\boldsymbol{a}(\boldsymbol{k})$ and $\boldsymbol{b}(\boldsymbol{k})$ define the lower and upper limit in each dimension (i.e., $\boldsymbol{a}(\boldsymbol{k})$ is a vector containing the lower limits of both dimensions), respectively, as follows: the average objective values in each dimension were transformed to subjective space and served as the lower or upper limits for a given category. For example, for the lower right quadrant in Fig. 2 (i.e., the Venak category) $\boldsymbol{a}(\boldsymbol{k})$ was $[\mu_{x1}, -Inf]$ and $\boldsymbol{b}(\boldsymbol{k})$ was $[Inf, \mu_{x2}]$. σ reflected variability in the perceptual distribution. That is, we make the assumption under this model that participants have a good understanding of the uncertainty (i.e., the variance–covariance matrix) of the perceptual representation. They use the noisily perceived value $\boldsymbol{x}_i$, which is used to calculate the cumulative probability that the observed stimulus belongs to any of the m categories.

3.2.3. Long-term memory gate

After feedback has been provided, a gate decides in the third stage whether the perceptual representation is advantageous for performance and stored as a memory representation or not. An important feature of the model is that representations are more likely to be stored in memory when they are helpful in the given task.

We varied the definition of helpfulness across two sampling algorithms. The first algorithm accepted every perceptual representation if it improved the probability of the currently presented stimulus to be classified correctly:

$$if \quad p(cat_k|\ \boldsymbol{x}_i)_t > p(cat_k|\ \boldsymbol{x}_i)_{t-1}, \tag{9}$$

with k being the index of the true category and t indexing the time point within the experiment. The second algorithm used a Metropolis–Hastings acceptance mechanism. A uniformly sampled value s between 0 and 1 was compared to the ratio of the category probability after feedback had been provided to the category probability before feedback has been provided. Whenever a sampled representation improves the probability of a stimulus to be classified correctly, it is accepted (i.e., the ratio is > 1). When a sample decreases this probability, it may still be accepted if s is smaller than the ratio. That is,

$$s \sim uniform(0, 1) \tag{10}$$

$$if \quad \mathrm{s} < \frac{p(cat_k|\ \boldsymbol{x}_i)_t}{p(cat_k|\ \boldsymbol{x}_i)_{t-1}} \tag{11}$$

The Metropolis–Hastings acceptance scheme was therefore more liberal in storing perceptual representations in memory than the acceptance scheme. For example, if a perceptual representation decreases the probability of a stimulus to be classified correctly, it may still be accepted with high probability. The bias due to storing helpful representations in memory is therefore smaller with Metropolis–Hastings than with acceptance sampling.

Accepting or rejecting samples in the sequential similarity task works similarly. We make the assumption that people use an internal representation of all stimuli to determine the identity of every presented stimulus. The internal representation is modeled as a normal distribution with the average mean values in subjective feature space and a standard deviation of .5. A perceived stimulus is represented as the cumulative density of the perceived value under all possible stimulus representations. The model accepts every perceptual representation if it is most likely under the true stimulus representation (improvement sampling). For the Metropolis–Hastings scheme, the model first calculates a normalized cumulative distribution of the likelihoods of the perceived stimulus given all true stimulus priors. It then again uniformly samples a value between 0 and 1, which lands on one stimulus portion of the cumulative distribution. The model then accepts the current representation under that stimulus prior. In essence, the model learns a more precise representation of every stimulus over time.

3.3. Model simulations

We conducted simulation studies with different model implementations in order to get qualitative predictions for representational change in the three different tasks. Two of the tasks were the category learning tasks with different category structures (i.e., ellipse or squared). The third task was the sequential comparison task. For the two category learning tasks, we varied the category learning model, the sampling scheme, and whether perceptual representations with values outside of the borders of the stimulus space exist. The latter was intended to explore potential effects that arise solely due to the edges of the feature space. For example, perceptual representations of stimuli close to the edges are more likely to be rejected solely due to the fact that a larger portion of their probability density is located outside of the feature space. For the sequential comparison task, we varied the sampling scheme and again whether samples outside of the feature space could be accepted or not. The standard deviation of the perceptual distribution was fixed to .5 for all simulations. We ran every simulation for 5000 trials.

A crucial question with regards to the task mapping part is what categorization strategy people actually use. The predictions of the task imprinting model we present in Fig. 6 are a mixture of the three strategies. In particular, we first ran the model for every strategy and every setting introduced in the previous paragraph. We then computed the weighted sum of the predictions from all strategies. To arrive at the model weights, we separately fit each category learning model to the categorization trials of every participant in all experiments using maximum likelihood parameter estimation (the modeling code is available on the OSF page). More specifically, we used the last 300 categorization trials to exclude initial learning from the model fitting. For Experiment 1, we used the naïvely Bayesian model as a prototype model, for Experiments 2–4, we used the multiplicative prototype model. In addition, we did not fit a rule-based model to the data of Experiment 1. The model comparison across all four experiments is displayed in Fig. 4. For simplicity, we only used participants for whom BIC and AIC were in agreement with each other. This yielded model weights of .824 and .176 for exemplar and prototype models in Experiment 1, respectively, and model weights of .706, .235, and .059 for exemplar, prototype, and rule-based models in Experiments 2–4, respectively.

We assume that participants have unbiased long-term memory representations of the stimuli after carrying out the initial task aimed at measuring representations, because their goal is to reproduce/estimate the identity of the stimulus as precisely as possible and because they do not receive feedback about their responses. As baseline models reflecting no or unbiased influence from long-term memory representations on perceptual representations before the start of the secondary task (see leftmost panels in Fig. 6), we trained the three category learning models to predict the true category labels using the average perceptual stimulus representations as input. We then ran a simulation and compared the stored memory representations after the 5000 trials with the average unbiased representations before the simulation. The mechanics of the model are visualized in Fig. 5.

The main prediction of the model is that unbiased responding changes into biased responding after category learning, but not after sequentially comparing stimuli. As can be seen in Fig. 6, representations of stimuli close to the decision boundaries are predicted to change more due to categorization practice than representations of stimuli further away from the boundaries. That is, the former representations are pushed away from the decision boundaries, which leads to representations from all categories but the residual category in Experiment 1 being assembled more closely to each other. The change is qualitatively similar across the two category learning structures. When translated to the average distance of representations to the associated category center, category learning decreases that distance for all but the residual category in Experiment 1 (i.e., Venak category in ellipse structure and any category in squared structure), but slightly increases that distance for the residual category in the ellipse setup (i.e., Bukil category). These average predictions are displayed in the middle row of Fig. 6. In addition to the average distance to the associated category center, we also calculated the average distance to the closest category boundary. A movement away from the closest boundary has previously been observed (Dubova & Goldstone, 2021) and is referred to as boundary aversion. The predictions with regards to boundary aversion essentially mirror those using distance to the associated category center. That is, distances for all but the residual category in Experiment 1 grow larger.

A second point to be mentioned is that the qualitative predictions of the model are relatively stable across different model implementations. The predictions do not depend on the used category learning model. That is, whether people use rules, prototypes, or exemplars to categorize stimuli affected the predictions of the model only marginally. Neither did the simulation results vary qualitatively as a function of the sampling algorithm. In general, though, the predicted effects were quantitatively larger for improvement sampling than for Metropolis–Hastings sampling.

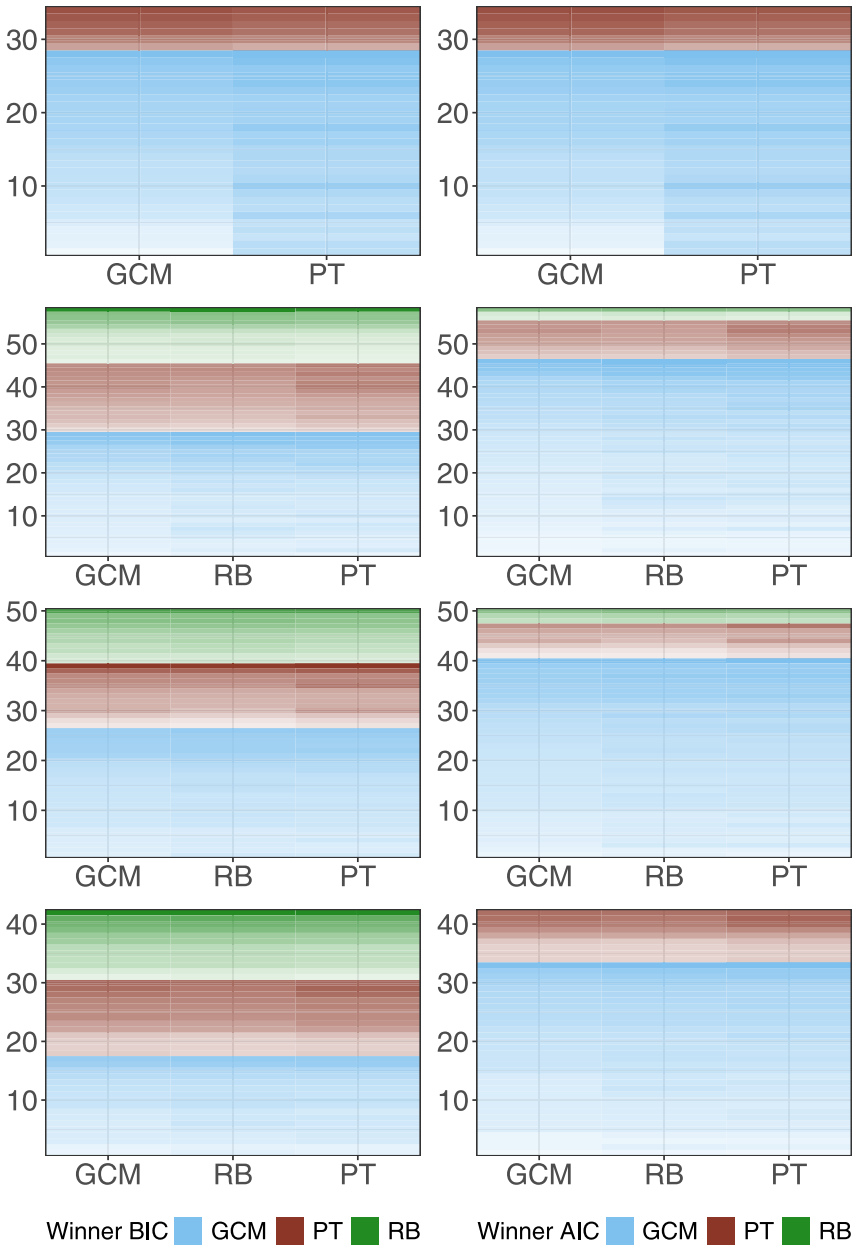


Fig. 4. Left and right columns compare the three models using BIC and AIC for model comparison, respectively. Rows 1 to 4 represent experiments 1 to 4, respectively. For a given subject (i.e., one row in one of the subplots), the color represents the winning model. The intensity of the colors represents the absolute value in ascending order (i.e., brighter means lower value and vice versa).

3.4. Locating the representational change

A general assumption we make here is that representations used to respond are a mixture of perceptual representations and memory representations (e.g., Hasantash & Afraz, 2020; Souza et al., 2021). Therefore, responses should be biased towards the memory representations. However, at this point we have to mention that other interpretations of the task imprinting model are possible. While we refer to the accumulated representations as stored long-term memory representations, it could equally well be argued that what changes over time is the perceptual representation itself. That is, the stored samples reflect a change in the perceptual representation itself over time and the predicted bias in responding is due to actual changes in the perceptual representation. Given the used tasks, it is hardly possible to differentiate between these different interpretations. Results from several decades of studies examining cognitive influences on perception cannot unambiguously be interpreted as evidence whether changed

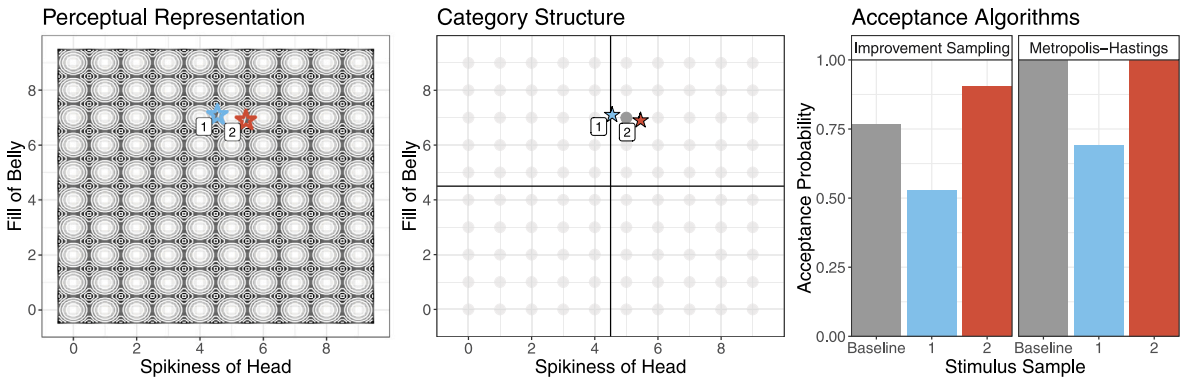


Fig. 5. Visualizes how different random samples from the perceptual distribution of the same physical stimulus lead to different probabilities of storing the stimulus in memory. In the hypothetical example, the model starts with unbiased representations of the stimuli and decides whether two samples (on two different trials) should be stored in memory according to the two sampling schemes. Left: In two different trials, two different values are randomly sampled from the perceptual distribution of the same stimulus. Middle: The stimuli vary in how much they are representative of the respective category. Right: The perceptual representations are stored in memory either when they increase the probability of a stimulus being categorized correctly (Improvement Sampling) or according to a storage probability (Metropolis–Hastings), which is calculated as the ratio of the probability of a stimulus being categorized correctly when the sample under consideration is added to the training data divided by the baseline probability without adding that sample.

responses are due to changed perceptual representations, due to changed long-term memory representations, or due to judgmental processes (e.g., Firestone & Scholl, 2016). Our current study does not present an exception with regards to that.

However, a main contribution of the current study is that we use tasks, which are more or less affected by judgmental processes. For example, the task to measure visual-perceptual short-term memory representations used in Experiments 1 and 2, is less affected by judgmental processes. Goldstone (1994) refers to these tasks as perceptual. In his definition, the tasks we use in Experiments 3 and 4 are associative. While associative tasks are affected by judgmental processes including task demands, perceptual tasks are affected by memory, but less or only marginally by judgmental processes. Therefore, observing evidence for task imprinting in Experiments 1 and 2 would be evidence for changed representations (perceptual or memorial, which of the two cannot be pinned down), changed responses in line with the predictions of the task imprinting model only in Experiments 3 and 4 would point towards judgmental processes.

4. Experiments

All four experiments were designed to test the qualitative prediction of the computational model that representations change differently according to task practice in the category learning task and in the sequential comparison task. The main difference across experiments was that we used tasks to measure representations that differ in the amount that strategic, judgmental processes play a role. In Experiments 1 and 2, we started by measuring predominantly visual-perceptual representations in a short-term memory task. In Experiment 3, we measured representations via similarity judgments of simultaneously presented stimulus pairs. In Experiment 4, participants were required to say how likely they thought two simultaneously presented stimuli belonged to the same category.

4.1. General method

All experiments used a two-groups between-subjects design. Participants were randomly assigned to one of two groups. Each group went through a set of three sequential stages. In the first stage, we attempted to get a baseline measurement of the representations of the full stimulus set. In the second stage, participants carried out the secondary task. One group carried out the category learning task as a secondary task, the other group carried out the sequential comparison task. In the third stage, we again measured representations of the full stimulus set. Essentially, task imprinting should be reflected in a detectable performance difference between stage three and stage one. Whereas the secondary tasks stayed the same across all four experiments (with one modification in the category structure, though), we varied the task to measure representations across the four experiments. Details are explained in the experiment-specific methods sections.

Participants for all experiments were recruited using the prolific platform (prolific.co). Fluent speakers of English aged between 18 and 35 with minimum approval rates of .9 and 1 for Experiments 1 and 2–4, respectively, and a minimal number of previous submissions of 5 and 20 for Experiments 1 and 2–4, respectively, were eligible to participate. By not further restricting participation in our experiments (e.g., to certain countries), we expect the results to generalize to young healthy adults in general. All participants agreed to take part in the study and were informed about the general purpose of the experiment. Experiments were performed in accordance with the relevant guidelines and regulations approved by the ethics committee of the University of Tuebingen (protocol nr. 701/2020BO, study title: Experimente zum Sequenz- und Belohnungslernen). Experiments were presented to participants using a combination of HTML, javaScript, and CSS with custom code. After a presentation of the instructions, participants were required

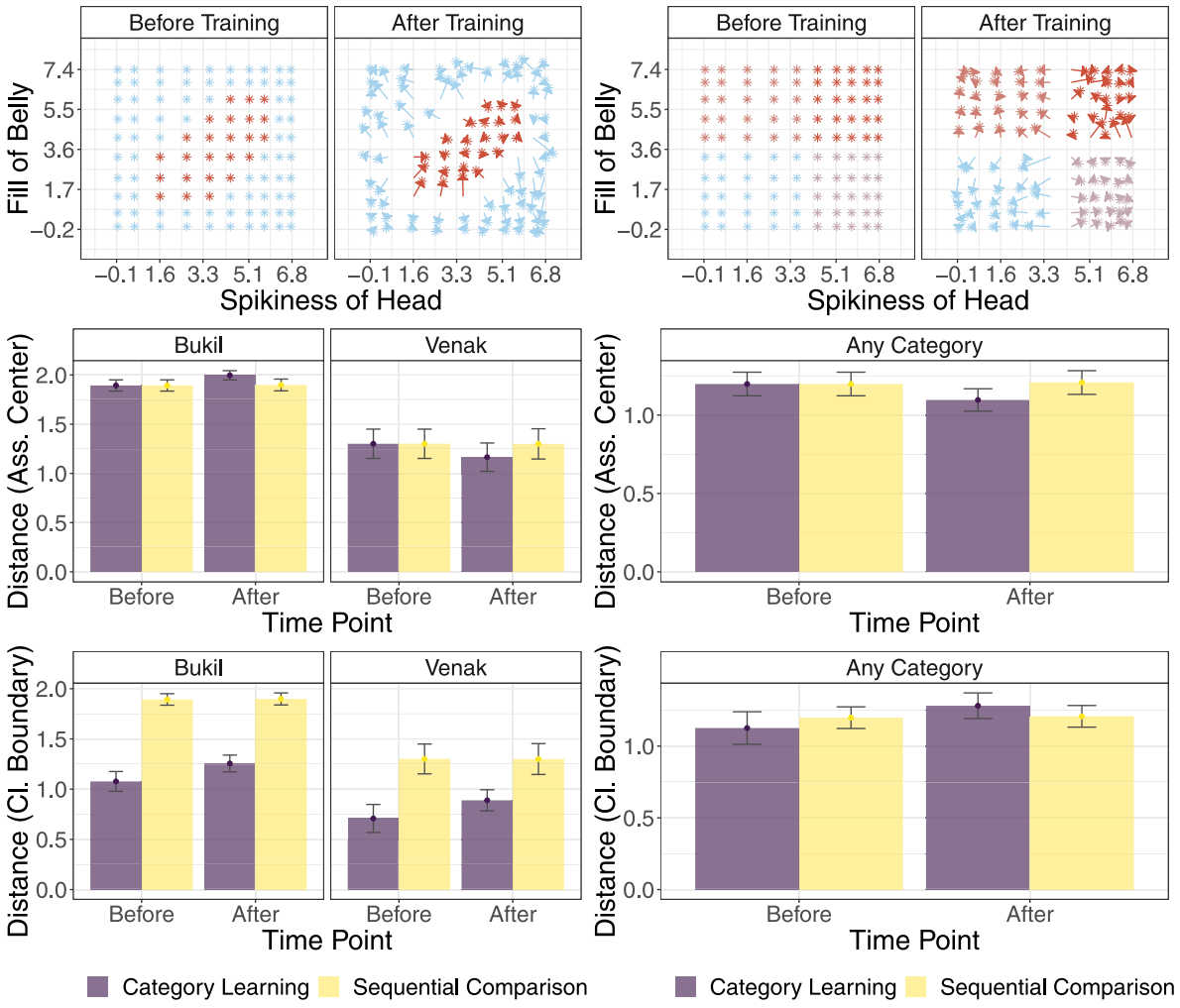


Fig. 6. Top row: Long-term memory representations of the stimuli in the two-dimensional feature space before and after category learning for the ellipse category structure (left) and the square category structure (right). Middle row: Average distance to the associated category center before and after the category learning task or the sequential comparison task, again plotted separately for the two categories. Bottom: Same as middle row, but for the distance to the closest category boundary. Because representational change is predicted to be similar for all four square categories, they are collapsed in the middle and lower right plot.

to complete a comprehension questionnaire. Only upon responding correctly to all questions, they could proceed to the main part of an experiment.

The stimuli were the same monsters in all four experiments, which differed from each other according to two continuous dimensions: the fill of their belly and the spikiness of their head (see Fig. 1 for two examples). Belly fill was defined as the radius of the circle inside the body, which resulted in a blue area of varying size. Head spikiness was defined as the radial distance between ten inner and ten outer vertices. These vertices were located between ten equally spaced segments of two concentric circles with different radii. In conjunction, they formed the ten spikes of the head when they were connected by a line. Values in both dimensions could be parametrically varied according to 100 steps (e.g., from an almost empty belly to a full belly). Participants were only exposed to a subset of 100 combinations in the two-dimensional feature space. We created these 100 stimuli in the following way: in each dimension we cut 9 steps from both ends. Then, we cut the remaining values from 10 to 91 into 9 equally sized segments of 9 steps. The 10 values from 10 to 91 were fully crossed across the two dimensions yielding 100 stimuli in the feature space.

We used Bayesian statistics for data analysis to overcome some of the shortcomings associated with frequentist statistics (Wagenmakers, 2007). In particular, we rely on the Bayes factor (BF) to quantify the evidence in favor of an effect of interest. The BF reflects the posterior odds of two models if their prior odds are .5. Kass and Raftery (1995) provide rough guidelines on how to interpret the strength of evidence of BFs: BFs ranging from 1–3.2 are not worth more than a bare mention, BFs between 3.2–10 provide substantial evidence, BFs between 10–100 are regarded as strong evidence, and BFs larger than 100 as decisive. To arrive at BFs for individual model parameters, we used the Savage–Dickey density ratio (e.g., Wetzels et al., 2009). This method provides

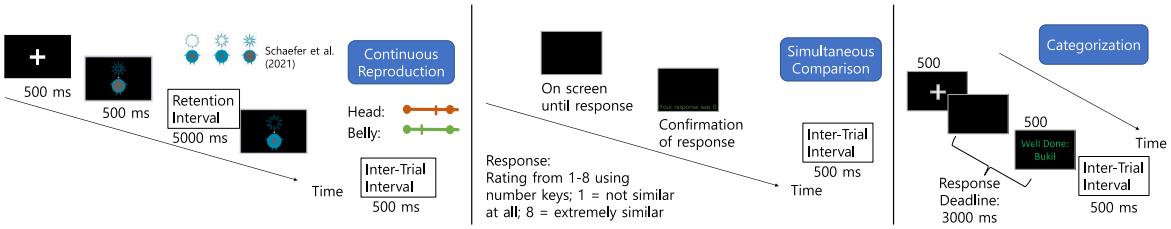


Fig. 7. Left: Procedure for the continuous reproduction task used in Experiments 1 and 2. The timing parameters were modified from Experiment 1. See text for details. Middle: Procedure for the simultaneous comparison task used in experiments 3 and 4. Right: Procedure for the category learning task used in all experiments. Note. The procedure for the sequential comparison task was the same as for the category learning task.

BFs for nested models. In particular, it compares a model allowing the parameter of interest to vary freely to a model fixing the parameter to the null model.

We report the structure of all used hierarchical regression models in the Appendix. Where necessary and possible (considering model convergence and run time), we first compared models varying in their random effects structure using the LOO method implemented in the loo R package (see Vehtari et al., 2017). We then used the winning model for inference about individual parameters (Matuschek et al., 2017). All Bayesian models were run in STAN (Carpenter et al., 2022) and accessed via the R programming language (R Core Team, 2022). The number of samples from the posterior was determined given ad-hoc considerations for every model to balance stability of the Bayes factor over several model runs and model run time. We additionally checked that all Rhat values were below 1.01. Details on the number of stored samples, burn-in samples, etc. can be found in the online accompanying material on the OSF page (see below). Note that we mean-centered all continuous predictors and simple-coded all categorical predictors (i.e., simple effect coding) before entering them into the models. These measures allow us (a) to interpret the fixed intercept as the grand mean, (b) to directly interpret parameters of categorical predictors as the effect of changing the predictor from the reference category (if there are more than two categories) to the category in question, and (c) to interpret all interactions entered into the model as cross-over interactions (Loftus, 1978).

4.1.0.1. Transparency and openness. All experimental scripts, raw data, and analysis scripts, as well as scripts for simulations with the computational model of representational change, are available on the following Open Science Framework webpage: https://osf.io/pgyr4/files. We pre-registered predictions, exclusion criteria, experimental designs, and more of Experiments 2–4 on the following OSF webpage: https://osf.io/pgyr4/registrations.

4.2. Experiment 1

In Experiment 1 we measured representations using a continuous reproduction task (Pertzov et al., 2013; Souza et al., 2014) intended to measure predominantly visual-perceptual representations of the stimuli.

4.2.1. Method

4.2.1.1. Participants. 84 participants (10 unknown, 45 women, 29 men) completed one session lasting approximately 90 min. They received a base payment of 9.80 GBP and an additional performance-dependent bonus of up to 5.20 GBP. 12 participants quit the experiment in between or provided incomplete data. We excluded 1 participant because their average distance from the true stimulus in the continuous reproduction task was larger than three times the standard deviation above the mean. We excluded 6 participants because their number of correct categorization responses (excluding the initial 40 trials, see below) was below the 99.9% percentile of a binomial distribution with probability .5. We excluded 3 additional participants because they restarted the experiment after they had already progressed through a substantial portion of the experiment. Thus, 62 participants remained in the experiment, 33 in the experimental group, 29 in the control group.

4.2.1.2. Materials and procedure. All 100 stimuli were presented once in the first session of the continuous reproduction task and once in the second session. Presentation order of the stimuli was randomized. For the whole duration of the continuous reproduction task, the background color of the window was white with the exception of an upward-facing rectangle in the center of the screen filled in black color. That setup was intended to increase the contrast between background and stimuli. A trial started by the presentation of a fixation cross in white color in front of the black rectangle for 500 ms. After that, a stimulus replaced the fixation cross and was displayed for 750 ms followed by a 2000 ms blank retention interval. Then, an average stimulus with a value of 50 on both feature dimensions was displayed on the screen. Participants could change the values in each dimension quasi-continuously (i.e., according to 100 steps) with two sliders that were presented below the stimulus. Once participants were satisfied with their response, they pressed a button to submit their response. The next trial started after an inter-trial interval of 500 ms. In the beginning of the experiment, participants completed two practice trials to get used to the procedure.

We used an ellipse category structure in Experiment 1 (see left panel of Fig. 2). We used this type of information-integration category structure because it assures that participants have to pay attention to both dimensions to categorize stimuli accurately (Ashby & Gott, 1988). The background coloring of the screen was the same in the category learning task as in the continuous

reproduction task. A trial started with presentation of a fixation cross for 500 ms followed by presentation of a stimulus. Participants were given a response window of 3000 ms to respond. When they responded within that period, a green message told them "Well done: Category X!" when they responded correctly, and a red message told them "Category would have been: Category X" when they responded incorrectly. When they responded after 3000 ms, a message appeared on screen, which indicated that they should respond faster. Responses were still collected, though. After participants had responded, the next stimulus was presented after an inter-trial interval of 500 ms. Responses were given by pressing digits 1 and 2 on the keyboard. Procedures of the set of tasks used in Experiment 1 are visualized in Fig. 7.

The category learning task lasted for 640 trials. The first 40 trials consisted only of examples from the ellipse category. In the instructions displayed to participants, we labeled the ellipse category as the target category called "Venak" and the residual category as the non-target category called "Bukil".[1] We presented 40 trials only from the target category to initially familiarize them with the more specific category. We created the set of 40 stimuli by first shuffling the stimuli from the ellipse category randomly, appending two sets of them and selecting the first 40 stimuli from that sequence. The remaining 600 stimuli were created in the same way, though, by appending several randomly shuffled sets of each category and selecting the first 300 stimuli from both resulting sequences. Participants were informed that the first 40 trials contained only examples from the Venak category, and they were informed after the 40 trials accordingly that "The initial block with target category examples is now over. In the following, your task is to categorize monsters from all categories [...]".

The procedure in the sequential comparison task was the same as in the category learning task. The 640 stimuli were created in the same way as above, though, without the constraint of having the same number of stimuli from different categories. Participants responded with digits 1–4 reflecting similarity judgments from not similar to very similar. The response window was the same as in the category learning task. Instead of a correct/incorrect feedback participants' responses were shadowed on the screen (e.g., "Your response was: 4") after every response. In both secondary tasks, stimuli were presented in four blocks with an equal number of trials. In between blocks, participants could recover for one minute or skip the break and continue immediately.

For every reported model in Experiment 1, we stored 5000 posterior samples from each of three chains after discarding 1000 warm-up samples.

4.2.2. Results

4.2.2.1. Category learning. Our main goal is to show that participants learned to discriminate between the categories over the 640 trials. As can be seen in the left panel of Fig. 8, the initial 40 examples from the ellipse category led to a head start for this category compared to the residual category. We binned the remaining 600 trials into blocks of 20 trials and analyzed the data with a hierarchical logistic regression. Category was added as a categorical predictor, trial block was entered as a linear predictor into the model. The head start in the ellipse category was reflected in decisive evidence for the main effect of category (BF > 100). However, accuracy in the residual category approached accuracy in the ellipse category over the course of the experiment, which was reflected in strong evidence for the category times trial interaction (BF = 28). The evidence for the main effect of trial was ambiguous (BF = 1.8). A potential point of concern is that participants were not able to discriminate between the categories sufficiently well on average, even after 640 trials of practice. Even though categorization accuracy was well above chance level, it plateaued at .73 for the residual category and at .76 for the ellipse category in the last block of 20 trials.

4.2.2.2. Sequential comparison. The main idea of the control condition was to expose participants to the same stimuli as in the experimental condition. If participants carried the task out as instructed, we would expect their similarity judgments to be predicted by the Euclidean distance between subsequently presented stimuli. To test that, we regressed participants' similarity ratings onto Euclidean distance as a linear predictor in a hierarchical model (see middle panel of Fig. 8). Subsequent stimuli were rated as more similar when their Euclidean distance in the two-dimensional feature space was smaller and vice versa, which was reflected in strong evidence for the main effect of distance (BF = 18).

4.2.2.3. Continuous reproduction. Our model predicted that distances to the category center should decrease for the ellipse category and increase for the residual category in the experimental group, but stay the same in the control group. That pattern is reflected in a three-way interaction between time point, category, and group on the distance measure. We broke it down to a two-way interaction on differences between distances after the secondary task and distances before the secondary task. Differences should become smaller for the ellipse category and larger for the residual category in the experimental group, but stay the same in the control group. The pattern is then reflected in the interaction between category and group. We tested the prediction in a hierarchical regression predicting difference scores using group and category as categorical predictors. The BF in favor of the interaction was .025 providing strong evidence against the hypothesis (i.e., a BF of 40 for the Null). There was also strong evidence against the main effects of category and group, respectively (BFs were .016 and .025, respectively). Together, the results suggest that representations did not become biased. The absence of bias is visible when inspecting the distribution of by-participant average movements towards the category center, which are scattered around zero in both groups for both categories (see left panel of Fig. 9). Given that there was substantial variability in category learning success, we explored whether it was positively related to movements towards category centers in the continuous reproduction task. We measured categorization success using two different approaches: once as final accuracy in the last block and once as the amount of learning calculated via the difference between final accuracy in the last

[1] In all current experiments, we did not randomize the labels across categories. Thus, there is a minor possibility that certain categories were learned better because participants memorized one label better than the other label.

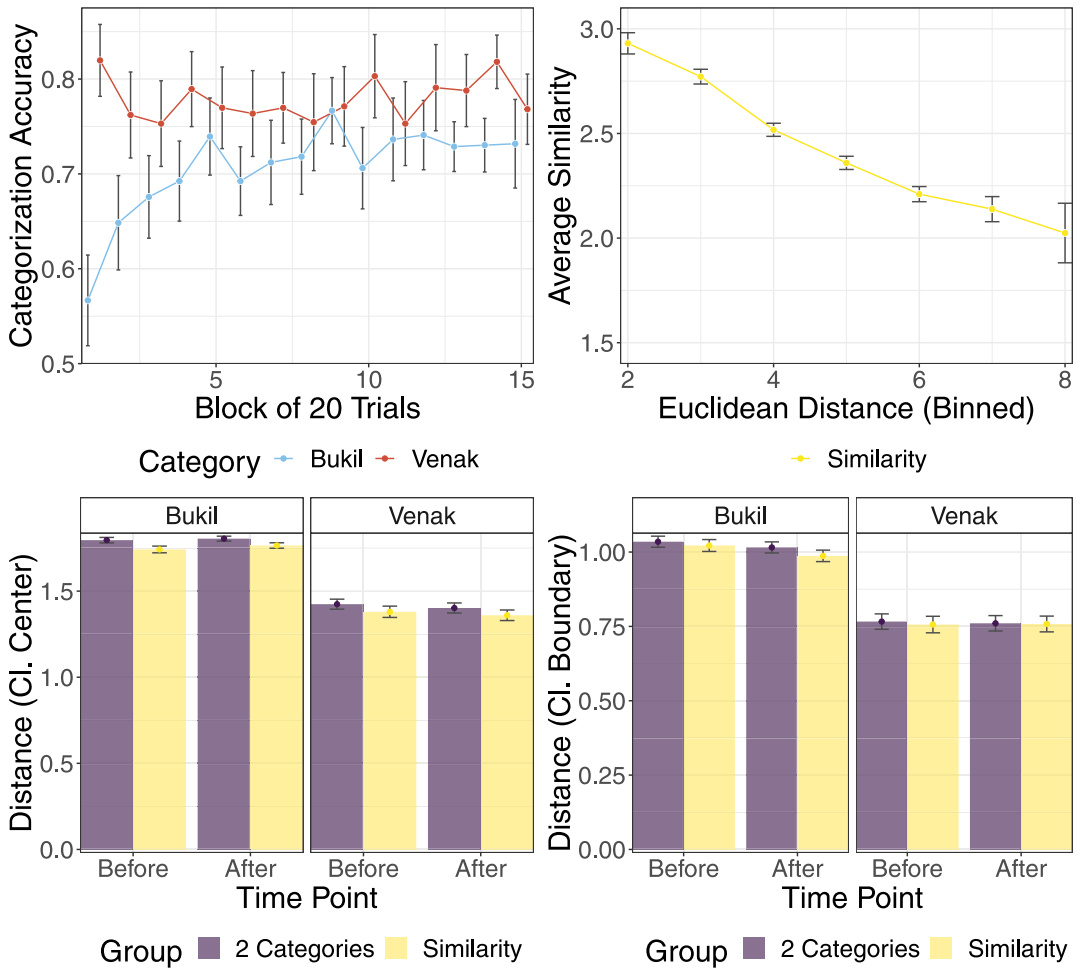


Fig. 8. Top left: Average categorization accuracy plotted per block of 20 trials separately for the two categories in Experiment 1. Note that the first 40 trials only including examples from the Venak category (i.e., from the ellipse category) are not presented. Top right: Average similarity rating plotted against Euclidean distance of subsequent stimuli in the sequential comparison task. Bottom left: Average distance to the category center only for stimuli of the ellipse category (i.e., using the same stimuli for both groups) in the continuous reproduction task before and after the secondary task. Bottom right: Same as bottom left, but with distance to the closest boundary plotted on the y axis. Note: Error bars represent 95% within-subjects confidence intervals.

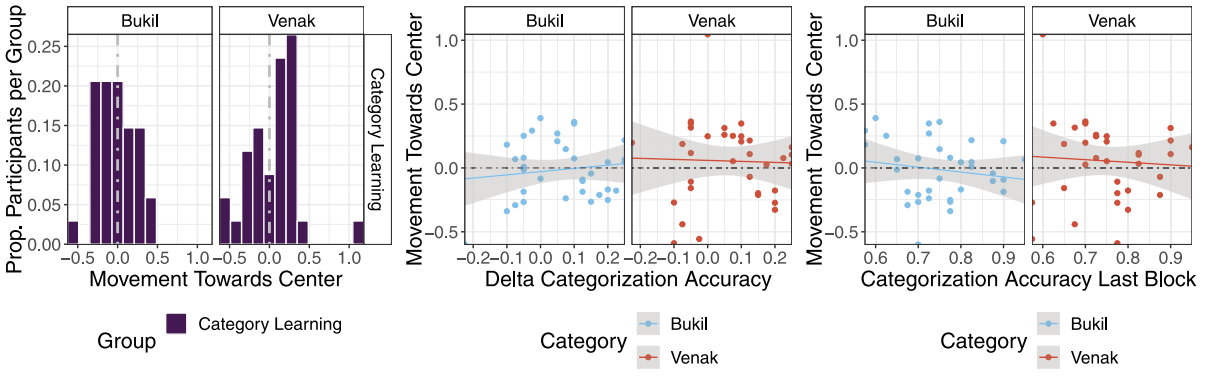


Fig. 9. Left: Histograms of by-participant averages of movements towards the category center plotted separately for the two categories (columns) and the two groups (rows) in Experiment 1. Middle: By-participant average movements towards the category center plotted against the average improvement in the category learning task. Right: By-participant average movements towards the category center plotted against the average final categorization accuracy in the category learning task.

block and initial accuracy in the first block. The models were implemented as fixed-effects linear regressions. Movements away from the category boundary would be reflected in an interaction between categorization success (linear predictor) and category (categorical predictor). That is, representations from the ellipse category should be pulled towards the category center more for successful learners than for less successful learners, whereas representations from the residual category should be pushed away from that center more for successful learners than for less successful learners. However, posterior distributions of the interaction in both models (using final categorization accuracy and the amount of learning) were centered close to zero and the respective BFs provided substantial evidence for the Null hypothesis (.30 and .27 for final accuracy and the amount of learning, respectively). The absence of these effects can also be seen by visually inspecting the middle and right panels of Fig. 9. Regression lines with a slope of zero are well within the shaded region of 95% frequentist confidence intervals.

In a similar analysis, we tested whether responses got pushed away from the decision boundary in the category learning group, but not in the sequential comparison group, reflecting boundary aversion. We tested boundary aversion as in the analysis above examining attraction to the prototype but replaced the dependent variable accordingly. It is reflected in the two-way interaction between category and group. The evidence, however, was strongly for the Null (BF = .02 in favor of the effect, hence a BF = 50 for the Null).

4.2.3. Discussion

Experiment 1 tested the idea that object representations, as measured in a continuous reproduction task previously used in the working-memory literature, become biased according to practice in a category learning task. Previous research showed that working-memory representations are affected by categorical knowledge (Donkin et al., 2015; Hasantash & Afraz, 2020; Huttenlocher et al., 1991; Souza & Skóra, 2017). The results provided, however, initial evidence against the idea of task imprinting. Representations did not change as a function of the secondary task at the group level. We additionally tested the idea that the amount of representational change is affected by the amount of learning in the categorization task. The results also provided evidence against that idea. Even participants who learned the categories well did not respond according to representations as predicted from the model.

One observation when analyzing individual differences in the amount of category learning was that a substantial proportion of the participants did not improve categorization performance at all over the course of approx. 550 trials (see middle panel of Fig. 9; several points are scattered around a delta of 0). That suggests that learning stopped relatively early for many participants. In addition, they may have learned relatively imprecise representations of the categories.

4.3. Conceptual changes in Experiments 2-4

A prerequisite for representational change in our model is that the category structure is learned sufficiently well. One possible reason for the absence of representational change in Experiment 1, therefore, is that participants did not learn to discriminate the two categories well enough from each other. Categorization accuracy plateaued at roughly .75 on average. We, therefore, applied two changes to the category learning task. First, we replaced the ellipse category structure with a potentially simpler squared category structure (see right panel of Fig. 2). Second, we motivated participants to perform well in the category learning task with two measures. Participants could only proceed to the third part of the experiment (i.e., the second measurement of representations) when they performed above one of two predefined performance thresholds. Additionally, they received a monetary reward when they surpassed one of the two thresholds. The calculation of the rewards is explained in Appendix. We also pre-registered the study designs, hypotheses, model predictions, analyses, exclusion criteria, and more in individual OSF pre-registrations for each experiment (https://osf.io/uvgc3).

4.4. Experiment 2

4.4.1. Method

4.4.1.1. Participants. 192 participants (23 unknown, 69 women, 100 men) completed one session lasting approximately 80 min. They received a base payment of 9 GBP and an additional performance-dependent bonus of up to 6.50 GBP. 71 participants did not reach the predefined performance criterion for their secondary task. We further excluded 2 participants because their average distance from the true stimulus in the continuous reproduction task was larger than three standard deviations above the mean. Thus, 118 participants remained in the experiment, 58 in the experimental group, 60 in the control group.

4.4.1.2. Materials and procedure. Materials and procedure of the continuous reproduction task were the same as in Experiment 1 with the following procedural changes: We reduced stimulus presentation duration to 500 ms and increased the duration of the retention interval to 5000 ms. We assumed that a potential influence from category knowledge on visual-perceptual representations is more likely with a shorter presentation duration and a longer retention interval (Donkin et al., 2015). We also changed the initial locations of the two sliders to reproduce the two feature values after stimulus presentation to random locations. Additionally, we applied several changes to the category learning task. Instead of two, there were now four categories defined as the four quadrants of the feature space (see the right panel in Fig. 2). We also reduced the number of trials in the secondary task (category learning and sequential comparison) to 400. Selection of the stimuli for the secondary task happened in the same way as in Experiment 1. Again, in both secondary tasks, stimuli were presented in four equally-sized blocks with the option of having a one-minute break. The initial 40 trials with only stimuli from the target category were dropped. Participants in both groups again were provided with immediate feedback about the correctness (category learning) or the identity (sequential comparison) of their just given response.

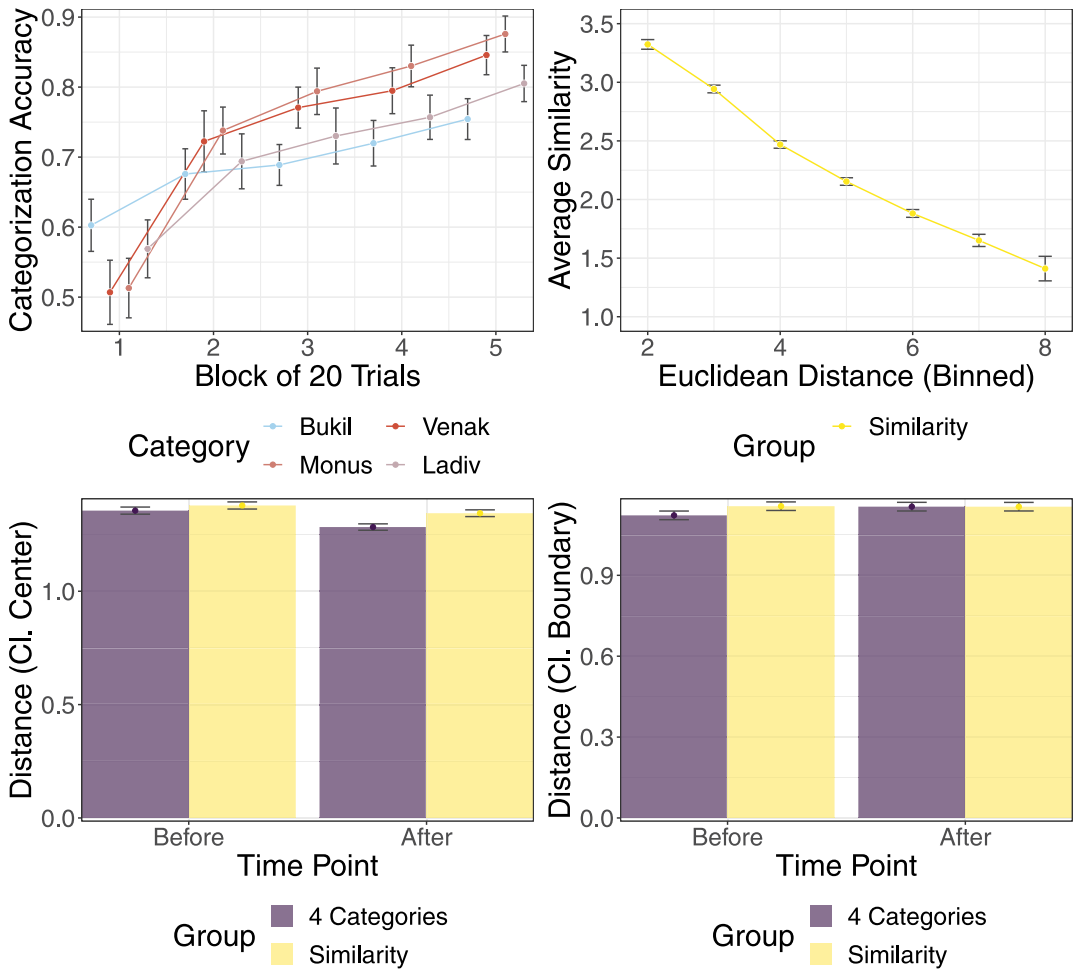


Fig. 10. Top left: Average categorization accuracy plotted per block of 20 trials separately for the four categories in Experiment 2. Top right: Average similarity rating plotted against Euclidean distance of subsequent stimuli in the sequential comparison task. Bottom left: Average distance to the associated category center in the continuous reproduction task before and after the secondary task. Bottom right: Same as bottom left, but with distance to the closest boundary plotted on the y axis. Note: Error bars represent 95% within-subjects confidence intervals. Distances in psychological space are square-root transformed in the lower two panels.

4.4.2. Results

For the category learning task, we dropped category and the category times trial interaction as predictors from the model. For the sequential comparison task, we used the same hierarchical model as in Experiment 1. Note that we used these two models for the same analyses in the remaining two experiments. Performance in the category learning task improved over blocks of 20 trials (BF > 100 for the main effect of block). The additional measures to increase performance in the category learning task also paid off. Final accuracy was at approximately .82 averaged over the four categories (see left panel of Fig. 10) with chance performance being at .25. Similar to Experiment 1, participants in the control group engaged well in the sequential comparison task. The effect of Euclidean distance on similarity ratings was decisive (BF > 100).

We analyzed the representational change in the continuous reproduction task with a hierarchical model predicting distances (square-root transformed) to the associated category center using group and time point as predictors (both entered as categorical predictors). The computational model predicts an interaction between time and group because the distances should become smaller for the category learning group, but not for the sequential comparison group. The results, however, provided strong evidence against that interaction (BF = .014) reiterating the null findings from Experiment 1. Visual inspection of the right panel of Fig. 10 suggests that the pattern of distances is at least qualitatively as predicted by the model. Therefore, we explored individual differences in the expected pattern in two further analyses.

We also tested the predicted representational change using distance to the closest boundary as the dependent variable. However, similarly as above, the evidence for the interaction between time point and group was strongly in favor of the Null (BF = .08).

The first analysis again tested whether success in the category learning task was positively related to movements towards the associated category center in the continuous reproduction task. Because movements are predicted to be qualitatively similar across

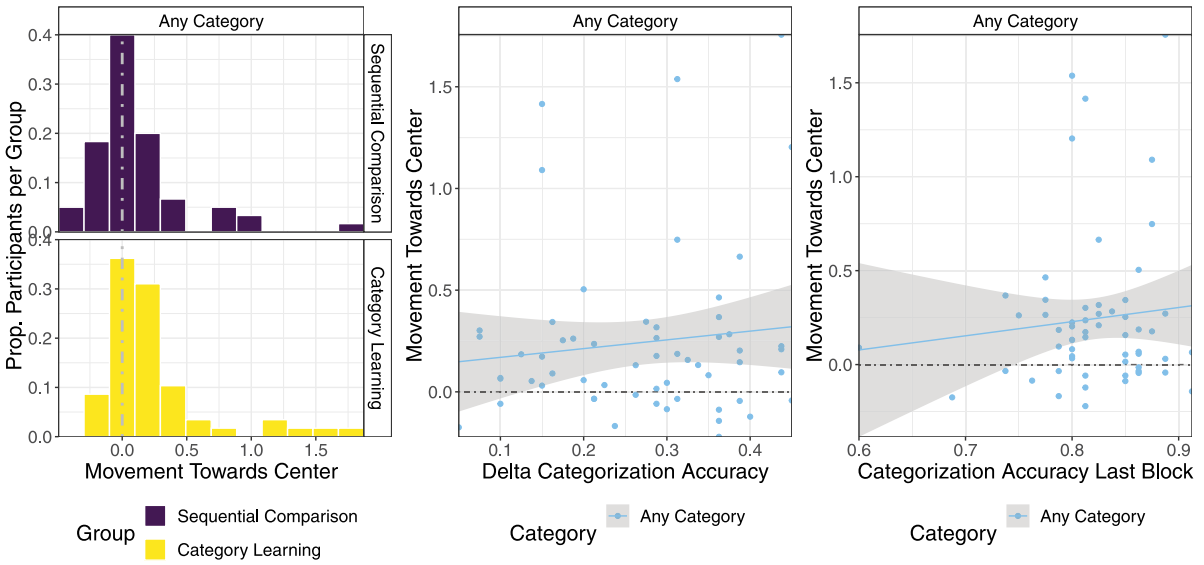


Fig. 11. Left: Histograms of by-participant averages of movements towards the category center plotted separately for the two groups (rows) in Experiment 2. Middle: By-participant average movements towards the category center plotted against the average improvement in the category learning task. Right: By-participant average movements towards the category center plotted against the average final categorization accuracy in the category learning task.

categories, success in the category learning task should be positively related to movements towards the associated category center in the continuous reproduction task. The analyses showed again no evidence for that idea (BF = .84 and BF = .81 for final accuracy and the amount of learning, respectively) (see Fig. 11).

The second analysis asked the question of whether continuous reproduction responses are the result of a mixture of stable responses and changed responses. We computed the before–after difference on the distance of the reproduction responses to the associated category center for every participant and every stimulus. The absence of a move towards the prototype would be reflected in a Gaussian distribution centered at zero. An average shift towards the prototype would be reflected in a shift of the mean of the Gaussian. We implemented this idea in a hierarchical model with a normal likelihood, in which the group means of the normal were allowed to vary across groups. Note that this model is similar to the one above predicting an interaction between group and time point on the raw differences. When only some responses were shifted, but not all, the shift would be reflected in a longer tail on the positive side of the distribution. We implemented the latter idea in a model, in which the likelihood of the data comes from an exGaussian distribution with the mean of the Gaussian component fixed at zero. For each group, we estimated a mean tau parameter (i.e., the exponential part), reflecting the average shift towards the category center for that group, with by-participant tau parameters hierarchically drawn from this group mean parameter. We freely estimated the sigma parameters (i.e., the standard deviation of the Gaussian part) for every participant (details about the model are available on the accompanying OSF page). The mean of the normal distribution centered at zero represents responses coming from unchanged representations. We formally compared the two introduced models, again with the LOO method (Vehtari et al., 2017), and the exGaussian model was slightly preferred over the shifted normal model (.51 vs. .49). Fig. 12 shows that both groups on average tended to shift their responses towards the category centers. The shift did however not differ between groups, as the 95% HDI of the difference between the group effects shows (see lower panel of that figure). Fig. 13 plots the by-participant MAPs against the empirical average moves towards the category centers surrounded by the two marginal histograms. It is visible that the pull towards category centers was driven by only a few participants in each group.

4.4.3. Combined analyses

4.4.3.1. Precision analysis. In a combined analysis of Experiment 1 and Experiment 2 we tested the idea that category learning affects perceptual representations of stimuli (e.g., Goldstone, 1994) over and above mere pre-exposure or pre-differentiation (e.g., Gibson and Walk, 1957). For example, Goldstone (1994) found that participants became more sensitized to dimensions that were categorization relevant. That finding was not restricted to comparisons between categories (i.e., acquired distinctiveness), but was also observed for comparisons within categories, providing evidence against the idea of acquired equivalence. The upshot of the current analysis is that we can contrast any potential performance increases only due to exposure to the stimuli in the sequential comparison task to any potential performance increases due to category learning.

For that reason, we coded responses in the continuous-reproduction task as differences from the true stimulus values on both feature dimensions. Note that for this analysis, we used the objective feature values to test whether responses became preciser. Descriptive statistics across all collected responses are shown in detail in Table 1 The resulting distributions before and after

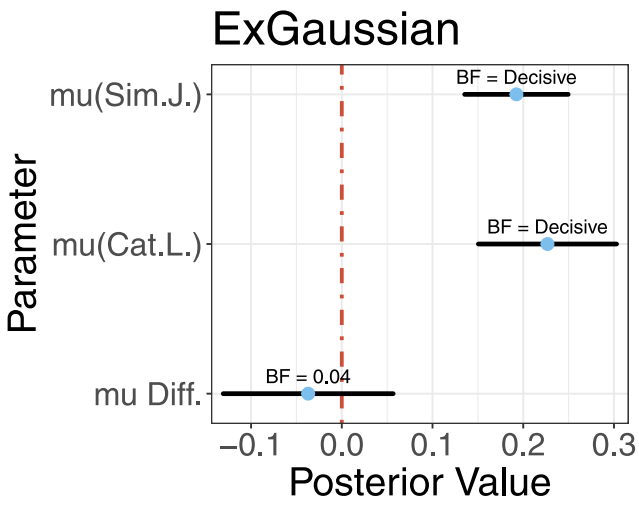


Fig. 12. Posterior MAPs, 95% HDIs, and BFs for the tau parameters on the group level in the exGaussian model.

Table 1
Descriptive statistics of all responses in the continuous reproduction task.

Secondary task	Timepoint	Mean		Std. Dev.		Correlation
		Head	Belly	Head	Belly	
Seq. comparison	Before	−0.46	1.04	17.13	15.61	0.04
Seq. comparison	After	−0.26	1.59	15.12	13.84	0.00
Category learning	Before	−0.33	1.39	16.04	15.51	0.03
Category learning	After	−0.12	0.88	13.42	12.34	0.02

secondary task practice are plotted in blue and red, respectively, in the left four panels of Fig. 15. In the right four panels, we plotted differences in the standard deviations (after secondary task–before secondary task) for the two groups and the two dimensions against distance to the closest boundary. It is visible that all differences were below 0, reflecting decreased standard deviations after the secondary task. We modeled the data as coming from a bivariate normal distribution with means of zero. We placed independent hierarchical regression models with random intercepts on the standard deviations. As predictors we used group, time point (both entered as categorical predictors), distance to the closest category boundary (entered as linear predictor), and all higher-order interactions between them. We also added experiment as a categorical predictor. Distance to the closest center was binned such that only stimuli immediately surrounding a category boundary were included in the first bin. Intermediate and large distances from the boundary were classified into the second and third bin, respectively. Pre-exposure to the stimuli predicts negative main effects of time point. If category learning increased perceptual sensitivity over and above pre-exposure, we would expect interactions between time point and group. Furthermore, if the standard deviation decreases more the further away from the category boundary, we would expect a negative three-way interaction between time point, group, and distance to the closest boundary (see Fig. 14).

The results show that standard deviations on both dimensions decreased over time (BF > 100). We additionally observed evidence for the time point times group interaction on the belly dimension (BF = 664), but indecisive evidence in the head dimension (BF = .46). The means of the posterior distributions of the interaction terms were negative in both two dimensions, though, with the 95% HDI excluding 0. That analysis is therefore consistent with the idea that learning to categorize stimuli affects perceptual sensitivity over and above exposure to these stimuli alone, particularly in the belly dimension. In addition, the analysis showed strong evidence that the standard deviation on the belly dimension, especially of stimuli further away from a decision boundary, decreased more due to category learning than due to sequential comparison (i.e., the three-way interaction). For completeness, we show the MAP estimates, 95% HDIs, and Bayes factors for all effects predicting standard deviations in the regression models in Fig. 15.

4.4.3.2. Movement towards the global average? Previous one-time-point designs showed that responses of participants tend to be shifted towards the global average of the feature space (Dubova & Goldstone, 2021). In an additional analysis, we tested whether this shift towards the global average changes after carrying out the secondary tasks. This analysis bears similarities with the attraction to the prototype analysis in Experiment 1. The main difference between the two was that we ignored category membership in this analysis. The results replicate the general pattern of responses being attracted towards the global average (all means are clearly above zero in the left two panels of Fig. 16). We then asked, whether this bias changes after carrying out the secondary task using a Bayesian ANOVA implemented in the BayesFactor package (Morey et al., 2024) in R with the anovaBF function specifying stimulus id as a random effect. In the right panel of Fig. 16 it is visible that the bias towards the global average became less strong (i.e., all difference scores are below 0), particularly for the category learning group (BF > 100) and in Experiment 2 (BF > 100). The evidence

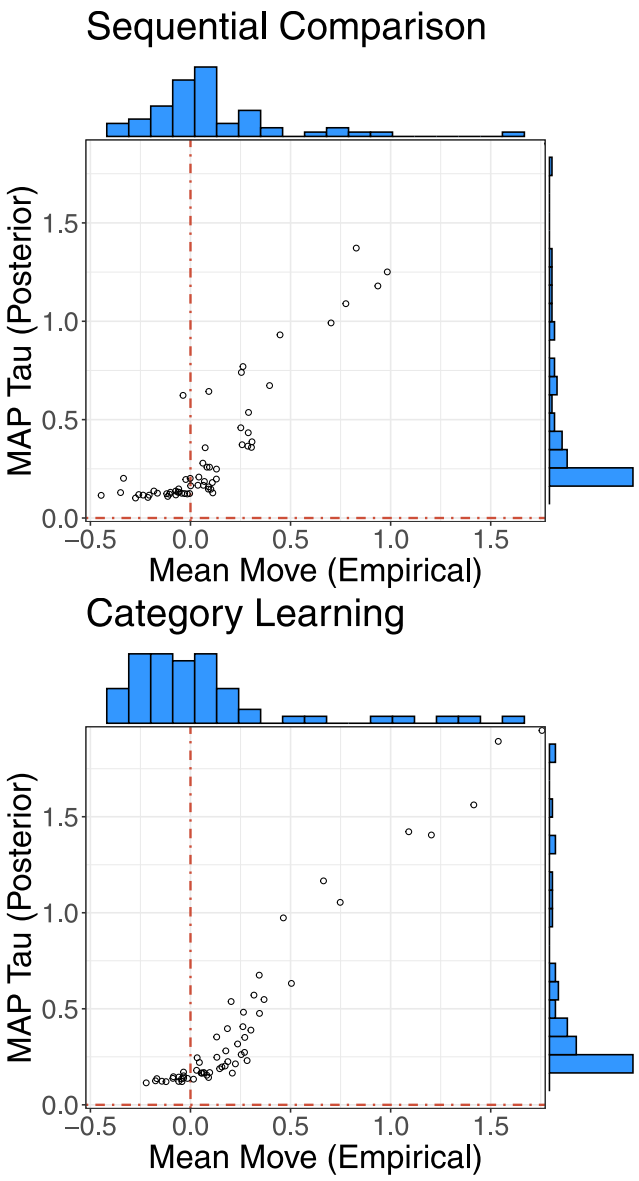


Fig. 13. By-participant MAPs plotted against mean empirical moves towards the category centers.

was ambiguous, whether the reduction in bias was larger for the category learning group than the sequential comparison group in Experiment 2 compared to Experiment 1 (BF = .80 for the interaction).

4.4.4. Discussion

The main goal of Experiment 2 was to increase learning in the categorization task to better test our main hypothesis. To achieve that, we made an effort to render category learning easier by replacing the ellipse category structure with the squared category structure and by additionally motivating participants with a monetary reward. Compared to Experiment 1, categorization accuracy in the last block increased by approx. .05 even though chance level was halved compared to Experiment 1 and participants had about 200 fewer trials to practice. We conclude, therefore, that the additional measures were successful. With regards to the main hypothesis of task imprinting, the pattern was qualitatively in agreement with the predictions on an average level (see right panel of Fig. 10). A formal analysis, though, again provided evidence against the predicted effect. In an attempt to quantify representational shifts and to evaluate individual differences therein, we fit a Bayesian exGaussian model to the data. The analysis showed that both groups only minimally tended to move responses closer to category centers, without any between-group differences. Most people showed none or a very small amount of response shifts towards category centers. Why would a subset of participants shift their responses to the category centers, even if they have not learned the categories in the first place? At this point, we can only speculate.

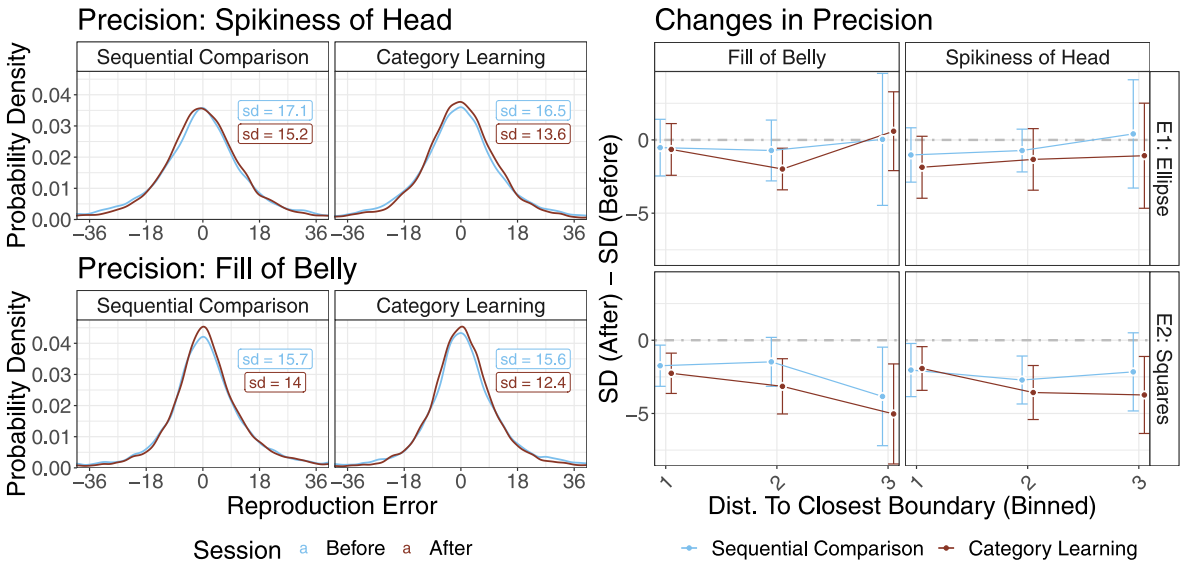


Fig. 14. Left panel: Marginal probability densities of the responses on the Head Spikiness dimension (upper row) and on the Belly Size dimension (lower row) in the continuous reproduction task before the secondary task (blue) and after the secondary task (red). The empirical standard deviations across all data points are shown as text labels within the respective panels. Right panel: Differences of the standard deviations (after) - standard deviations (before) plotted against distance to the closest boundary separately for the two groups. All descriptive statistics of the responses in the continuous reproduction task are shown in Table 1 below.

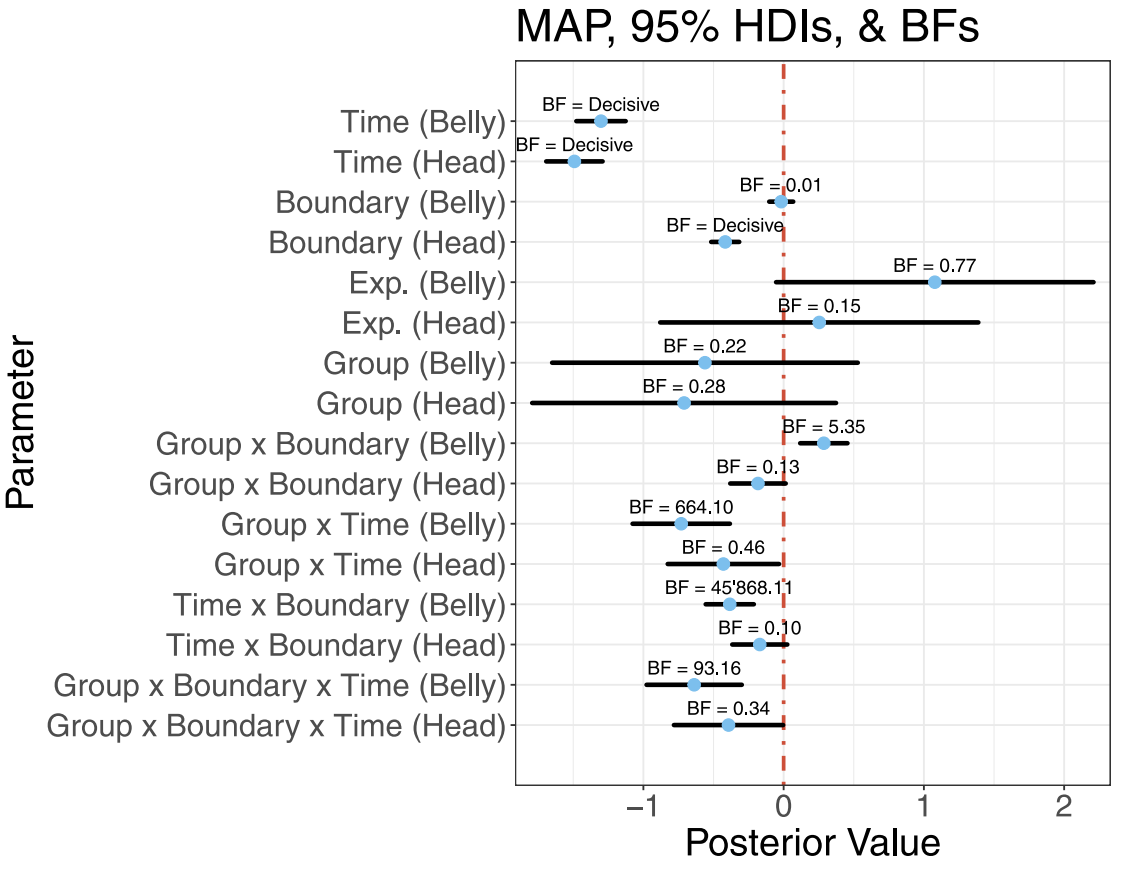


Fig. 15. Maximum a posteriori estimates (MAP), 95% highest density intervals (HDIs) and Bayes factors shown for all effects entered into the regression predicting standard deviations of the reproduction responses. Note. The intercept parameters are not displayed.

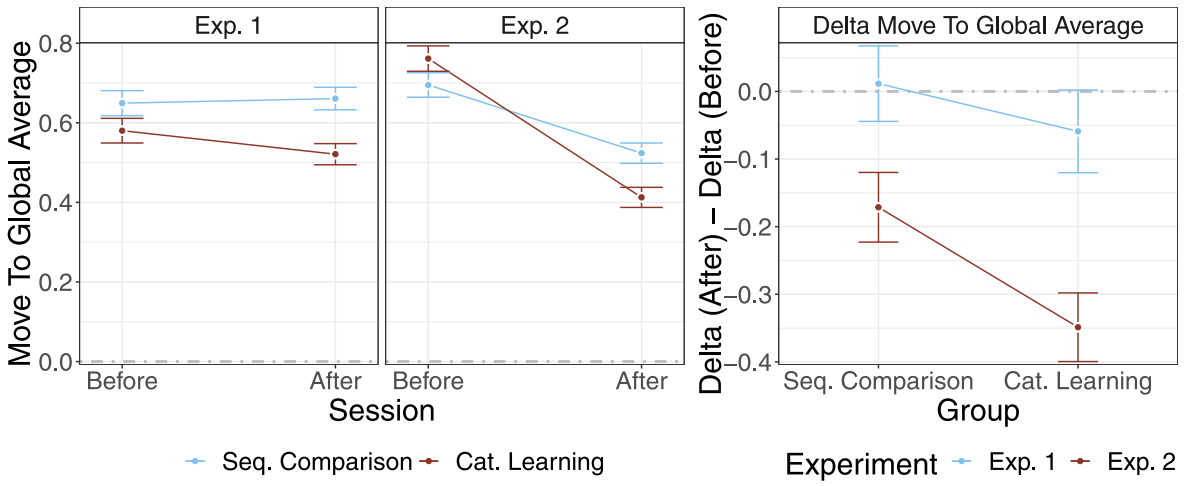


Fig. 16. Left panels: Movements towards the global average. Data points were obtained by first averaging across all responses for a given stimulus and then averaging across stimuli. Right panel: Same data as in the left panel, but differences after–before were calculated. Note. Error bars represent 95% confidence intervals.

If we assume that participants strategically tended to partition the feature dimensions into five distinct values (i.e., lo, lo-med, . . .) over the time of category learning, for example because of being annoyed to respond very precisely, responding accordingly would show similar signs as attraction to the prototypes.

A combined analysis of the data from Experiment 1 and Experiment 2 asked the question of whether improved perceptual sensitivity is only an effect of being exposed to a set of stimuli or whether learning to categorize these stimuli additionally increases it. The results partially supported the latter idea. Even though both groups responded more precisely in the continuous reproduction task after the secondary task than before, the increase in precision was larger for the category learning group than for the control group in the belly dimension. The results were inconclusive in the head dimension, but the MAP and the 95% HDI of the group effect were at least consistent with this prediction. The pattern is consistent with an account of increased perceptual precision due to category learning. The pattern is also consistent with an account postulating the storage of preciser exemplars in long-term memory during category learning, which eventually assist to reproduce stimuli from degraded working-memory representations, for example via redintegration (Lewandowsky, 1999).

In an additional analysis we tested whether bias towards the global average increased after carrying out the secondary tasks. The results are not in line with Bayesian models of perception or rate–distortion theory (Sims, 2016), which predict that this bias should increase, especially for the sequential comparison group. The basic idea is that the estimate for the average stimulus (i.e., the stimulus prior) becomes preciser after more exposure to the stimulus set. The results, however, are in line with the finding that responses generally become preciser with more exposure and category learning. This increased precision alleviates the general tendency of shifting responses towards the global average.

Two patterns do not align with previously reported results. First, the absence of a movement towards category centers or away from category boundaries does not conceptually replicate Hasantash and Afraz (2020)'s finding that continuous reproductions of colors tended to be pulled towards idiosyncratic color categories. Whereas these authors used pre-existing, idiosyncratic categories, we let participants learn the categories during the experiment. Therefore, it could be that the bias arises when categories are highly over-learned, such as colors, and does not reflect an adaptation to task-specific goals as suggested by the task imprinting model. It could also be that such a bias only arises for even lower level stimuli than used in the current experiment (i.e., colors, orientations). Second, the reduced standard deviation of reproduction responses, particularly in the category learning group, correlated negatively with distance from the category boundary. This does not replicate Goldstone (1994), who observed the opposite pattern of an increased precision of categorization-relevant separable dimensions especially close to the boundaries. Because we presented 100 different items with 10 feature values on each dimension as compared to four feature values in Goldstone (1994)'s study, the category boundary might have been less salient in our study. That is, whereas 75% of the items were adjacent to a boundary in the latter study, in our study only 34% in Experiment 1 and 36% in Experiments 2-4 were adjacent to a category boundary.

To summarize, in combination with Experiment 1, we interpret the set of results as in disagreement with the main qualitative prediction from the task imprinting model that responses become biased after category learning but not after sequential comparisons. A result in line with representational change, though, was that the precision of continuous reproduction responses increased, particularly after category learning. Whether this change was on the level of perceptual representations or on the level of preciser long-term memory representations, can however not be answered with the current data.

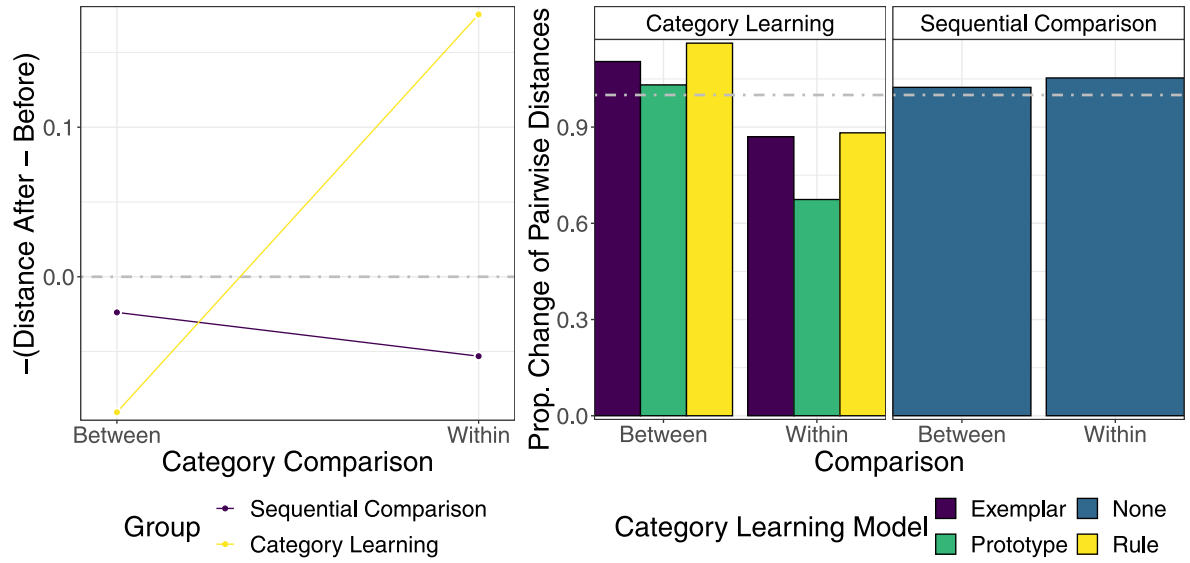


Fig. 17. Model predictions for the two secondary tasks. The qualitative pattern looked similar for the different model instantiations. Here, we present exemplary predictions for the model also accepting samples outside the pre-defined feature space using an acceptance sampling scheme. In the *left* panel, we grouped pairwise comparisons between two stimuli into pairs coming from the same category ("Within") and pairs coming from different categories ("Between"). We then plotted the difference between the distance after carrying out the simulation and the distance using the initial settings before running the simulation. To relate to similarities, we plot the negative value of this difference. In the *right* panel, we plot the proportion of change of the pairwise differences. That is, values below 1 reflect smaller distances after running the simulation, values above 1 reflect larger distances after running the simulation.

4.5. Experiment 3

In Experiment 3, we tested the idea that task imprinting reflects a judgmental process (Goldstone et al., 2001). It could be that people preferentially store certain representations in the category-learning task in long-term memory, but then are unaffected by them in the continuous reproduction task. Why would this be the case? This pattern could emerge if the stored representations are only used as conceptual representations. So a participant may decide not to use them in the reproduction task, because they are not helpful. We therefore replaced the reproduction task, which measures predominantly perceptual representations, even though conceptual influences have been shown (Donkin et al., 2015; Souza & Skóra, 2017), with a simultaneous comparison task. While it is assumed that similarity ratings in a simultaneous comparison task measure a perceptual component to some degree (Medin et al., 1993), Barsalou (1982) and Medin et al. (1993) argued that the information used in pairwise similarity judgments is affected by the context, in which a pair of stimuli appears. Similarly, Holt et al. (2014) showed that there is a substantial conceptual component in these ratings over and above the mere perceptual component. Hebart et al. (2020) argued that representations can only be measured in a context-free manner in an odd-one out task if the same pair of stimuli is presented over a large number of trials in the contexts of many different other stimuli. The learned category of an object may therefore serve as additional information in the simultaneous comparison task being used strategically via potentially activated representations stored in long-term memory.

The predictions from the computational model, therefore, do not differ from the previous two experiments on the level of individual stimuli. However, they were transformed into distances between pairs of stimuli before and after the secondary task. The 100 representations before (see left panels of Fig. 6) and after the secondary task (see middle panels of Fig. 6) were therefore each transformed into 10'000 pairwise distances between stimuli. We further classified the pairs into those coming from the same quadrant in the feature space and those coming from different quadrants (side-by-side quadrants and cross quadrants were collapsed). The predictions can be inspected in Fig. 17. Differences between model variants were again minor with the main expected pattern being distances between representations from the same category becoming smaller and distances between representations from different categories becoming larger. Whereas these representations were directly used to respond in the continuous reproduction task in Experiments 1 and 2, here, they are used as the input for the similarity judgments between two stimuli in the simultaneous comparison task.

4.5.1. Method

4.5.1.1. Participants. 160 participants (11 unknown, 77 women, 72 men) took part in one session lasting approximately 45 min. They received a base payment of 5 GBP and an additional performance-dependent bonus of up to 2.50 GBP. 57 participants did not reach the predefined performance criterion for their secondary task. We re-coded the responses of 8 participants, because they rated stimuli as more similar in the simultaneous comparison task when the distance between the stimuli increased. Then, based on this correlation between similarity judgments and Euclidean distance we excluded 2 further participants because their correlation

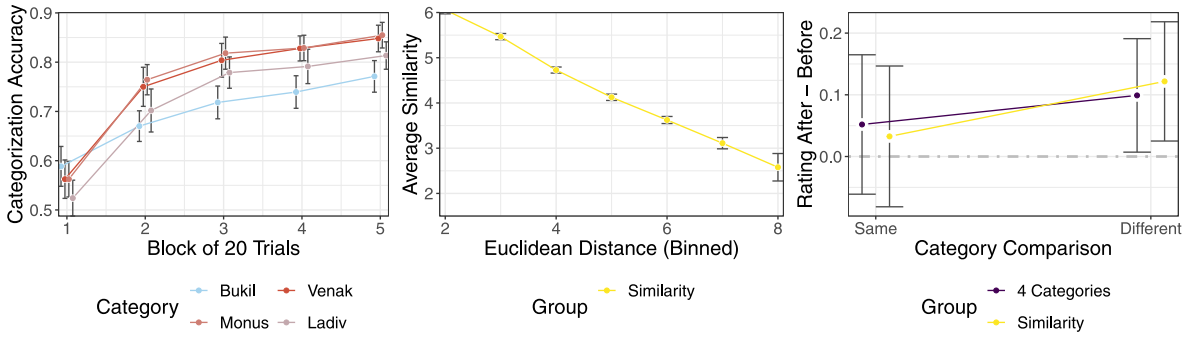


Fig. 18. Left: Average categorization accuracy plotted per block of 20 trials separately for the two categories in Experiment 3. Middle: Average similarity rating plotted against Euclidean distance of subsequent stimuli in the sequential comparison task. Right: Difference between the similarity ratings given to the same pairs of stimuli after vs. before. Note: Error bars represent 95% within-subjects confidence intervals.

was larger than three standard deviations above the mean correlation. We excluded 3 additional participants because they restarted the experiment after they had already progressed through a substantial portion of the experiment. Thus, 98 participants remained in the experiment, 52 in the experimental group, 46 in the control group.

4.5.1.2. Materials and procedure. We changed the task to measure representations from a continuous reproduction task to a pairwise simultaneous comparison task. In every trial, participants observed two monsters presented side by side and were instructed to "rate the monsters by how much you think they look similar to each other" using digits 1–8 (1: not similar at all, 8: very similar) on the keyboard. Stimuli were presented on screen until participants gave a response, which was followed by a message indicating their given response within a 500 ms inter-trial interval. Presentation of the next pair followed immediately. Similar to Experiments 1–2, the two stimuli were presented on black backgrounds on a white screen. We adapted the response scale in the sequential comparison task of the control group to match the scale in the simultaneous comparison task in order to avoid response confusions between the two conceptually similar tasks. Otherwise, the procedure of the sequential comparison task and the procedure of the category learning task were exactly the same as the ones in Experiment 2.

In order to avoid ceiling and floor effects in similarity ratings, we strategically sampled pairs for every participant according to the following procedure: First, we binned Euclidean distances of stimulus pairs into five buckets with thresholds 0, 30, 50, 70, 90, and ∞. We then assured that the distances of pairs to be rated for comparisons within and between quadrants of the feature space were on average not too small and too large, respectively. We did so by assuring that pairs sampled for every participant satisfied the following distributions over distance bins: [.5, .5, 0, 0, 0], [.1, .4, .3, .2, 0], and [0, .2, .3, .4, .1] for comparisons within the same quadrant (x4), between quadrants touching side by side (x4) and between quadrants touching only with their corners (x2), respectively. We randomly sampled 10 pairs for every quadrant comparison for every participant. The same participant saw the same 100 pairs twice, once before the secondary task, and once after the secondary task.

4.6. Results

The evidence was again decisive that participants in the experimental group improved in the category learning task over blocks (BF for main effect of block > 100). Similarly, those in the control group performed as expected in the sequential similarity task and rated more distant stimuli as more dissimilar and vice versa (BF for main effect of binned distance > 100) (see Fig. 18).

We then tested whether the similarity judgments changed due to category learning. It can be seen in Fig. 17 that the model predicts an interaction between category comparison (i.e., stimuli from the same category or stimuli from different categories) and group. Pairwise distances of stimuli within the same category should decrease and pairwise distances of stimuli from different categories should increase for the experimental group, but not for the control group. That pattern would be reflected in an interaction between group and category comparison on difference scores. Because people responded on a similarity scale, the direction of the effect should be mirrored. We tested this prediction in a hierarchical model predicting differences in similarity judgments using group and category comparison as categorical predictors. The BF in favor of the group times category comparison interaction was .05, therefore providing strong evidence for the absence of the effect.

In an attempt to account for individual differences in the mapping of psychological distance to similarity ratings, we related psychological distance to similarity ratings with a negatively accelerated exponential function. Note that this model is not the full GCM model, but approximates the similarity judgments using only the exponential function from the GCM. Individual attention weights for the two dimensions were drawn from a group distribution. Additionally, we placed a hierarchical regression model with random by-participant intercepts on the c parameter in a stan model. Besides the random intercept, there were fixed main effects for group, category comparison, and time point (all entered as categorical predictors). We also added all higher-order interactions of these variables into the model. Task imprinting would be reflected in an increased c parameter for comparisons between categories (i.e., less generalization) and a decreased c parameter for comparisons within categories (i.e., more generalization) for

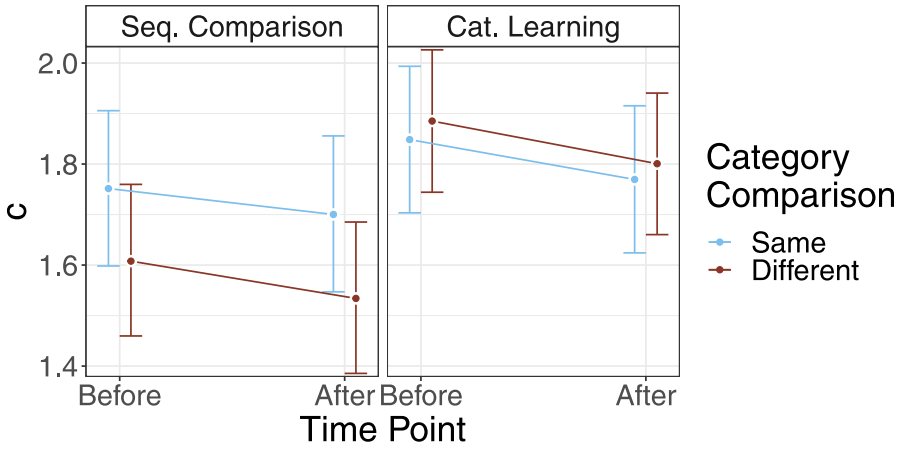


Fig. 19. Estimated c parameters for the model relating physical distance to psychological similarity for the two groups in Experiment 3.

the experimental group, but not for the control group. This prediction would be reflected in evidence in favor of the three-way interaction. The results (see Fig. 19) showed, however, strong evidence for the absence of such an effect (BF = 0.02, i.e., BF = 50 for the Null). There was also decisive evidence for the group times category comparison interaction, likely driven by the higher baseline in the c parameter in the category learning group for the comparison of items from different categories. There is a possibility that the sampled pairs from different categories for the category learning group were on average closer to each other in the feature space than in the sequential comparison group. However, as we were only interested in the change after carrying out the secondary task, we do not further interpret that effect.

4.7. Discussion

In Experiment 3, we tested the idea that task imprinting reflects a strategic, judgmental process. We therefore measured representations with a simultaneous comparison task. It has been argued that pairwise similarity ratings are affected by context (Barsalou, 1982; Hebart et al., 2020; Medin et al., 1993). We therefore hypothesized that preferentially stored representations during the category learning task may be used as additional information to respond in that task. The results were clear-cut. Even though participants learned to categorize stimuli well and responded as expected in the sequential comparison task, their responses did not change as a function of the secondary task. The conclusion stayed the same when we accounted for individual differences in the mapping from stimulus space to psychological space with a cognitive model.

4.8. Experiment 4

Whereas the previous experiments showed evidence against the idea that responses in the continuous reproduction task and similarity judgments in the simultaneous comparison task change according to the idea of task imprinting, we modeled the task to measure representations in Experiment 4 as even closer to the category learning task. The idea was similar as in Experiment 3, in particular to create a context, in which category membership becomes even more salient. Therefore, we emphasized category membership even more in the instructions. To achieve that, we asked people to judge how likely two simultaneously presented stimuli belong to the same category. Evidence for task imprinting in such a task in combination with the previous null findings would suggest that changed responding is mostly strategic Goldstone et al. (2001), that the representational change is tightly linked to the practiced task, and that it generalizes only very narrowly and in a task-specific fashion.

4.8.1. Method

4.8.1.1. Participants. 113 participants (5 unknown, 40 women, 68 men) took part in one session lasting approximately 45 min. They received a base payment of 5 GBP and an additional performance-dependent bonus of up to 2.50 GBP. 26 participants did not reach the predefined performance criterion for their secondary task. We further excluded 3 participants because the correlation between their similarity judgments and Euclidean distance was larger than three standard deviations above the mean correlation. We excluded 2 additional participants because they restarted the experiment after they had already progressed through a substantial portion of the experiment. Thus, 82 participants remained in the experiment, 42 in the experimental group, 40 in the control group.

4.8.1.2. Materials and procedure. Materials and procedure were the same as in Experiment 3 with the following exception in the instruction: in the simultaneous comparison task participants were instructed to respond by how likely they thought that the two stimuli belonged to the same category. Note that the control group was never exposed to any categories, but was still asked about how likely they thought that two stimuli may belong to the same category.

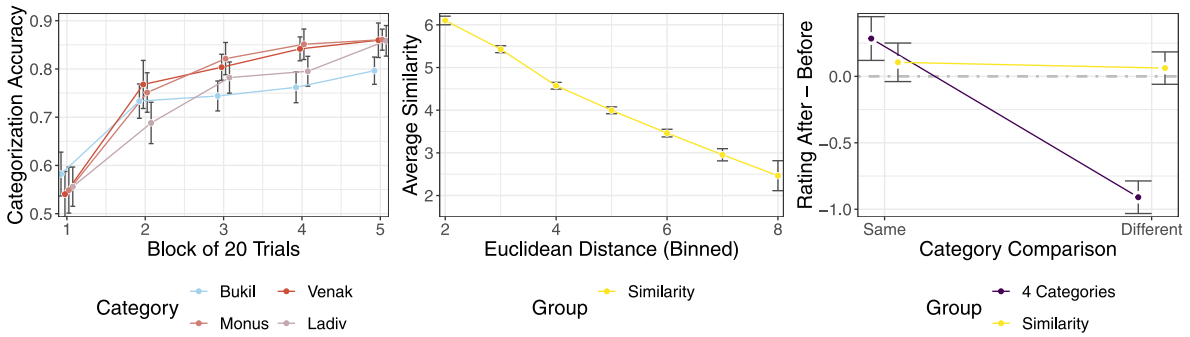


Fig. 20. Left: Average categorization accuracy plotted per block of 20 trials separately for the two categories in Experiment 4. Middle: Average similarity rating plotted against Euclidean distance of subsequent stimuli in the sequential comparison task. Right: Difference between the similarity ratings given to the same pairs of stimuli after vs. before. Note: Error bars represent 95% within-subjects confidence intervals.

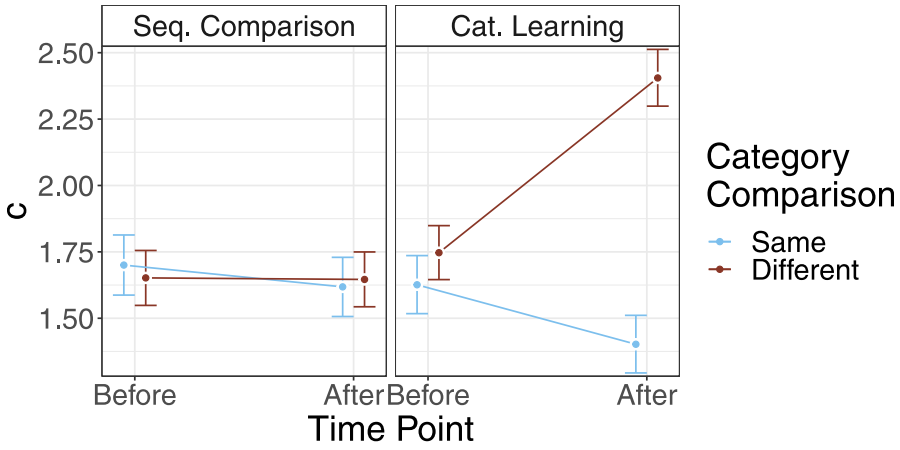


Fig. 21. Estimated c parameters for the model relating psychological distance to similarity for the two groups in Experiment 4.

4.8.2. Results

As in the three previous experiments, (a) categorization performance improved substantially over blocks (BF > 100) and (b) similarity ratings in the sequential comparison task were a negative linear function of Euclidean distance between subsequent stimuli (BF > 100). Model predictions for the simultaneous comparison task were the same as for Experiment 3; the pattern may even be quantitatively amplified, because participants were explicitly instructed to rate the likelihood of category membership. The results confirmed the predictions. Similarity ratings for stimuli from the same category slightly increased whereas similarity ratings for stimuli from different categories decreased for the experimental group, but not for the control group (BF = 88 for the category comparison times group interaction; see right panel in Fig. 20). The conclusion remained the same when accounting for individual differences in the mapping between psychological space and similarity ratings in the c parameter. Generalization between stimuli was only modulated by the secondary task in the experimental group, but not in the control group. The BF in favor of the three-way interaction effect was decisive (BF > 100, see Fig. 21).

4.8.3. Discussion

In Experiment 4, we modeled the context of the task measuring representations even closer to the category learning task than in Experiment 3. The idea was that representations stored in long-term memory during the category learning task are more likely to be used with an even more overlapping task context. The results provided evidence that participants' responses changed as predicted from task imprinting. Whereas stimuli were perceived as similar across the two groups before the secondary task, only category learning but not sequential comparisons changed their interpretation. After the secondary task, participants in the category learning group rated stimuli from different categories as less likely to come from the same category and tended to rate stimuli from the same category as more likely to come from the same category. Responses in the control group remained unchanged. The conclusion remained the same when we accounted for individual differences in the mapping from physical space to psychological space with a computational model.

5. General discussion

Research from several areas suggests that object representations are affected by knowledge about the same objects accumulated in a different task. In the current paper, we proposed a new mechanism of this type of representational change relying on cross-task talk, task imprinting. The assumed mechanism was inspired by previous research about the particular phenomenon that short-term memory responses are affected by previously learned categorical information (e.g., Donkin et al., 2015; Hasantash & Afraz, 2020; Huttenlocher et al., 1991; Souza & Skóra, 2017). We implemented the idea of task imprinting in a computational model based on principles of perceptual variability or noisy stimulus encoding akin to population coding (Ashby & Lee, 1993; Ashby & Townsend, 1986; Bays, 2014; Pouget et al., 2000) and preferential storage of helpful information in long-term memory.

We tested the model's predictions in a series of four experiments. The main qualitative prediction of the model was that representations of stimuli, which are assumed to be a mixture of perceptual and memory representations, close to a decision boundary drift away from that boundary towards the category center. For all categories the shift should direct away from the category boundary and for all but the residual category in Experiment 1, the shift should direct towards the respective category center. Only Experiment 4 showed signatures of task imprinting. In that experiment the task context of the task aimed at measuring representations was intended to be very similar to the category learning task. The continuous reproduction task, which we used in Experiments 1 and 2, was modeled according to previous research that found influence from categorical knowledge on behavior (Donkin et al., 2015; Hasantash & Afraz, 2020; Huttenlocher et al., 1991; Souza & Skóra, 2017). Despite achieving acceptable levels of category learning in Experiment 2, the results showed evidence against task imprinting. We assumed that representations, which have been stored in long-term memory during the category learning task as predicted by our model, would assist as additional information to make the pairwise similarity judgments in Experiment 3. The reasoning was that pairwise similarity ratings are more likely to be context-specific (Barsalou, 1982; Hebart et al., 2020; Holt et al., 2014; Medin et al., 1993) and therefore activate categorical information from long-term memory. However, similarity judgments in Experiment 3 were left completely unaffected by learned category knowledge. To test whether the results depended on the specific method we used to measure representational change, we also analyzed representational change with the representational behavioral similarity analysis (see Appendix, Karagoz et al., 2022). The results, though, showed the same pattern as the analyses reported in the main part of the manuscript.

The results overall do not support the idea that representations of objects are biased by helpful representations in another task. They point more towards an account that participants strategically change their responses given task demands. For example, a potential alternative explanation is the strategic judgment bias account (Goldstone et al., 2001). That account assumes that categorical information serves as an additional cue, which is added to the computation when determining the similarity between two presented stimuli. As our proposed account of task imprinting, however, it would have predicted an effect on similarity judgments in Experiment 3. By additionally assuming that only tasks sufficiently similar to a category learning task prompt the strategical use of categorical cues, the account could be made consistent with the current results.

Similarly, our results are partially in line with the literature on categorical perception. On the one hand, the combined analysis of Experiment 1 and Experiment 2 showed that category learning increased perceptual sensitivity as assessed via the continuous reproduction task over and above mere pre-exposure to the same stimuli, a similar effect already shown previously by Goldstone (1994). On the other hand, we did not observe any differential effects from category learning on similarity judgments for stimulus pairs between categories and stimulus pairs within categories. Such an effect would be predicted if stimuli from the same category were harder to differentiate (i.e., acquired equivalence) and stimuli from different categories easier to differentiate (i.e., acquired distinctiveness).

Previous studies showed that responses in short-term memory tasks are affected by categorical knowledge (Donkin et al., 2015; Hasantash & Afraz, 2020; Huttenlocher et al., 1991; Souza & Skóra, 2017). Experiments 1 and 2 could not experimentally induce a tendency to use categorical representations over and above visual-perceptual representations. The observation in Experiment 2, however, that some participants used stereotypical responses, as opposed to categorical representations, adds a new flavor to some of the previous studies. For example, responses shifted to stereotypical angles and locations on a circle in Huttenlocher et al. (1991) might also be partially due to stereotypical responding. That is, participants just preferentially responded at a few stereotypical locations from the whole space of possible responses.

5.1. Comparison to other models

Is the main qualitative prediction of the task imprinting model unique? Even though we believe that the mechanics and the implementation we propose are novel, previously introduced Bayesian models of cognition, connectionist models, and rate–distortion models can predict the same pattern of shifted representations towards category centers or away from category boundaries. For example, in the Category Adjustment Model by Huttenlocher et al. (2000) stimulus estimates are constructed by combining an imprecise stimulus representation (i.e., the likelihood of the data) with a category representation (i.e., the prior), which is represented by the central category value. This bears clear similarities with our implementation, and the predictions for the currently used designs are the same. However, the current predictions from the task imprinting model are more general, because they do not depend on the assumption that categories are represented as prototypes. The same prediction can be derived from connectionist models. For example, in the model by Harnad (1995), which is sequentially trained to auto-associate and categorize one-dimensional stimuli, attraction towards category averages evolves as a natural side-effect due to general principles of how neural networks learn. Similar to the task imprinting model, the emergence of categorical perception can be described as adaptive to the dual-task demands of the auto-association task and the categorization task. Finally, rate–distortion models, which bear similarities to Bayesian models, but add the assumption of a capacity-limited channel to transmit information from input to output, can make a similar prediction (e.g., Sims, 2016). The larger the capacity limitation, the more these models predict an assimilation towards prototypes (e.g., the mean value), because a precise representation of an individual stimulus becomes too costly.

5.2. Broader context

In line with previous work (e.g., Firestone & Scholl, 2016; Goldstone, 1994), we made the point that behavioral responses most often cannot unambiguously be interpreted as purely driven by perceptual representations and/or long-term memory representations. The latter can moreover be stimulus specific (i.e., Nosofsky, 1986) or categorical/conceptual. The strong version of the task imprinting model predicted a change on both levels, such that perceptual representations become biased by long-term knowledge in order to become a better conceptual or categorical learner. The result that people only changed their responses accordingly in a situation, in which they seemed to be able to infer the correct response, shows that people can decide to use their knowledge in some situations, but not in others. This aligns with work on the development of knowledge structures, which has shown that outdated knowledge and newly acquired knowledge, for example about the solar system, coexist together in the same individuals (Shtulman & Harrington, 2016). Similarly, Vosniadou and Brewer (1992) showed that when children are asked to change their mental model of the world, for example when they learn that the earth is a sphere, they integrate the new knowledge in such a way that they change as few of their presuppositions as possible. The current findings add to the understanding of how several concepts about the same objects develop and coexist together mentally.

Our results also shed new light on findings from neuroscience that showed that representations as measured via activation patterns in the hippocampus change according to concept learning (Mack et al., 2016; Theves et al., 2019, 2020, 2021). Especially the finding that representations become preciser, but not biased as predicted by the task imprinting model seems relevant. Correlations between neural activation and concept space after category learning are therefore less likely to be attributed to changed perceptual or memory representations of individual objects than to learned representations of the categories. We suspect therefore that the adaptation reflects the changed meaning of the stimulus in the context of the category learning task. Given the current results, we would predict that presenting the same stimulus in different task contexts would elicit different patterns of activation in the hippocampus once the different tasks have been learned sufficiently. That could be tested, for example, by presenting a random pre-cue indicating one of two category structures to be used for an upcoming categorization. The current results also add to the discussion about what mechanisms repetition suppression reflects. In their review, Barron and colleagues listed four mechanisms of repetition suppression proposed in the literature (Barron et al., 2016). All of them emphasize the identity of the stimulus to be crucial for a repetition suppression effect to emerge. Our findings permit an interpretation of repetition suppression effects not purely based on stimulus properties. We propose that repetition suppression measures adaptation to a context-specific representation of a stimulus. The same stimulus may not give rise to a repetition suppression effect if it was used in a different task.

5.3. Conclusion

In the current work, we introduced a computational model predicting a mechanism of representational change called task imprinting. We tested the main qualitative prediction of the model in a series of four experiments. Based on the results, we can exclude that task imprinting is a general, context-free phenomenon biasing perceptual and memory representations of individual objects. The results suggest much more that perceptual representations are strategically re-interpreted within a given task context using additional knowledge stored in long-term memory — for example representations that were helpful in another task. People are more likely to respond with signatures of task imprinting the closer the task context of the task used to measure representations is to the task context in which a particular representation was helpful.

Funding

This work was funded by the Max–Planck institute for biological Cybernetics, Germany.

CRediT authorship contribution statement

Mirko Thalmann: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Theo A.J. Schäfer:** Writing – original draft, Software, Resources, Investigation, Conceptualization. **Stephanie Theves:** Project administration. **Christian F. Doeller:** Writing – original draft, Project administration. **Eric Schulz:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Data availability

Experimental scripts, modeling scripts, and analysis scripts are available on the following osf webpage: https://osf.io/pgyr4/files/ Pre-registrations of Experiments 2–4 are available on the following osf webpage: https://osf.io/pgyr4/registrations/.

Acknowledgments

We thank Alex Kipnis for valuable input on notational issues, Marcel Binz, René Schlegelmilch, and two anonymous reviewers for valuable feedback on a previous version of this manuscript.

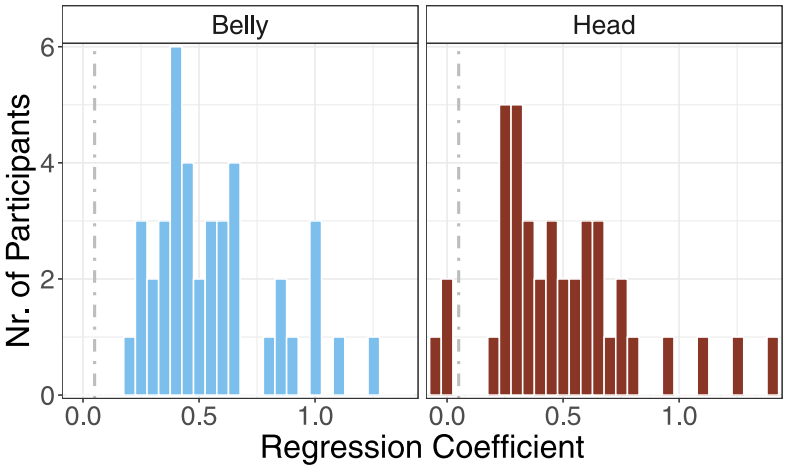


Fig. 22. Psychophysical Scaling Experiment. By-participant beta coefficients relating object properties to psychological values (see lower panel Fig. 3), which we derived using the maximum-likelihood difference scaling approach (Maloney & Yang, 2003).

Appendix

A.1. Psychophysical scaling experiment

A.1.1. Participants and methods

51 participants (5 unknown, 17 women, 29 men) took part in one session lasting approximately 25 min in an online study on prolific. They received a fixed payment of 4 GBP for their participation. The experiment was programmed using the software jsPsych (De Leeuw, 2015). Due to technical problems, we had to exclude the data of 11 participants, because a substantial amount of the data was not saved for them (<80% of their data were saved). We further excluded 3 participants whose responses were not significantly affected by the intensity of the feature values (see head dimension in right panel of Fig. 22 and read further below for the specific metric used).

In every trial of the task, participants were presented with three stimuli. They had to decide whether a target stimulus presented centrally in the bottom of the screen was more similar to one of two reference stimuli presented in the top left and top right of the screen. In half of total 364 trials, the value of only one dimension was varied in 14 linearly spaced steps of equal size. The value of the other dimension was set to the average objective value. From all possible combinations of stimuli, we selected those, in which the target stimulus had a value in between those of the two comparison stimuli. We randomly divided the 364 possible triplets in each dimension into two sets a and b of each 182 triplets. Every participant observed 182 triplets in each domain resulting in the total 364 trials. Half of the participants observed sets a, half of the participants observed sets b. The two dimensions were presented in blocked order. The order of presenting the sets (i.e., first head spikiness dimension or first belly fill dimension) was counterbalanced across participants. In each dimension, participants received 5 initial practice trials.

A.1.2. Results

For every participant, we applied the MLDS function in the mlds package in R (Knoblauch & Maloney, 2023) to the by-trial responses, separately for every participant. We then fit a linear regression predicting each participant's values on the psychological scale using the objective values as predictor. The distribution of coefficients is plotted in Fig. 22. We excluded participants whose responses were not significantly ($p < .05$) related to the objective properties (i.e., participants whose coefficient was below the gray dashed line in Fig. 22). We then averaged the psychological values across participants in each dimension and fully crossed the resulting values. These average values are used throughout the study, and they are visible in Fig. 2.

A.2. Hierarchical structure of Bayesian models

Random effects were random with regards to participants except for the analysis modeling attraction towards the global average, in which we modeled stimulus id as a random effect.

- **Experiment 1**
  - Category Learning: random intercept, random slope trial, fixed slope category, fixed slope category *x* trial (note. models with random slopes on category or random slopes on category and category times trial did not converge)

  - Sequential Comparison: random intercept, random slope distance
  - Movement (i.e., differences between distances before and after the secondary task) Towards Category Center/Closest Boundary: random intercept, random slope category, fixed slopes number of categories and number of categories times category
  - Individual Differences in Movements: simple linear regression without random effects.

- **Experiment 2**

  - Category Learning: random intercept, random slope trial
  - Sequential Comparison same as for Experiment 1
  - Movement Towards Category Center/Closest Boundary (square-root transformed): random intercept, random slope time point, fixed slopes number of categories and number of categories times time point; note. this model fit better than a model with random-intercept only.
  - Individual Differences in Movements: simple linear regression without random effects.
  - Move Model "Normal Shift": Random intercept
  - exGaussian: random intercepts for sigma and tau parameters

- Combined Analysis

  - Precision Model: fixed mean value to zero on both dimensions; hierarchical regression on standard deviations with random intercept only and fixed effects of boundary distance, number of categories, time point, all higher order interactions between the latter three predictors, and experiment
  - Global Average Model: random intercept across items, fixed effects of group, experiment, and the group times experiment interaction

- **Experiment 3**

  - Category Learning and Sequential Comparison same as for Experiment 2
  - Differences in Similarity Judgments: random intercept and random slope category comparison, fixed slopes number of categories and number of categories times category comparison; note. this model fit better than a model with random-intercept (without random slope category comparison) only.

- **Experiment 4**

  - All three models were specified in the same way as in Experiment 3

A.3. Monetary rewards

- Continuous reproduction: We calculated the average deviation from a participant across all trials. When the average deviation was below 5, a participant got the largest bonus. When the average deviation was above 51, a participant did not get any bonus. Average deviations between 5 and 51 were linearly scaled from the maximum bonus to no bonus.
- Category learning: The proportion of correct responses was multiplied by the maximum bonus.
- Sequential comparison: All participants were paid the same average bonus.
- Simultaneous comparison in Experiment 3 and Experiment 4: The correlation between Euclidean distance and participants responses was multiplied with the maximum bonus.

The maximum rewards in

- Experiment 1 were 2.60 £ and 2.60 £ for the continuous reproduction task and the category learning task, respectively. We paid 1.30 £ for the sequential similarity task by default.
- Experiment 2 were 3.25 £ and 3.25 £ for the continuous reproduction task and the category learning task, respectively. We paid 2.85 £ for the sequential similarity task by default.
- Experiment 3 and Experiment 4 were 1.25 £ and 1.25 £ for the simultaneous comparison task and the category learning task, respectively. We paid 0.85 £ for the sequential similarity task by default.

A.4. Bayesian group-level posterior estimates

For the sake of completeness, here we report the MAPs, 95% HDIs, and the BFs for all group-level effects of the Bayesian models used in Experiments 1–4 (see Figs. 23–28).

A.5. Behavioral representational similarity analysis

We also analyzed task imprinting using the behavioral similarity approach proposed by Karagoz et al. (2022), called behavioral representational similarity analysis (bRSA). Therefore, we calculated the change in distance for every pair of stimuli when comparing

Fig. 23. Experiment 1.

responses before the secondary task and responses after the secondary task. In Experiments 1 and 2, we calculated the mean Euclidean distances between all pairs of stimuli before and after the secondary task. In Experiments 3 and 4, we calculated the mean difference in simultaneous comparison judgments for all the rated pairs, also before and after the secondary task. We then calculated the difference for every pair after vs. before. We proceeded in the same way for the model predictions and then correlated differences according to the model with the observed differences in the four Experiments. Correlations between model matrices and responses should be positive for the experimental group in Experiments 1 and 2 because predictions and data reflect distances, but negative for Experiments 3 and 4 because data are similarity judgments, but predictions are distances.

bRSA Experiment 1

r = .03 (sequential comparison), .01 (category learning) (see Fig. 29).

bRSA Experiment 2

r = .01 (sequential comparison), .07 (category learning) (see Fig. 30).

The RSA plots for E3 and E4 show some empty cells, as not all 10k pairs were observed by the participants.

bRSA Experiment 3

r = .03 (sequential comparison), .01 (category learning) (see Fig. 31).

bRSA Experiment 4

r = .00 (sequential comparison), -.17 (category learning) (see Fig. 32).

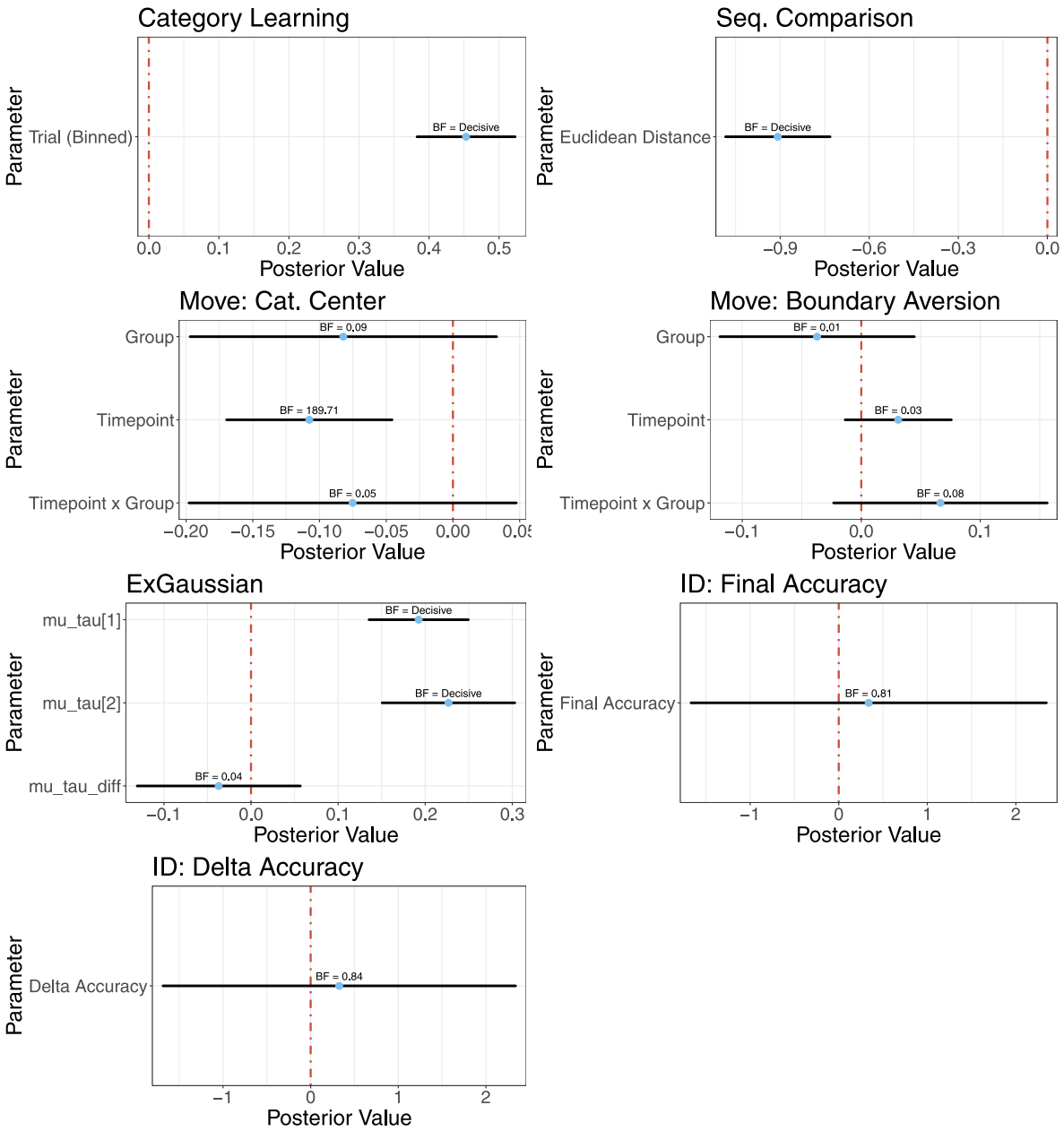


Fig. 24. Experiment 2.

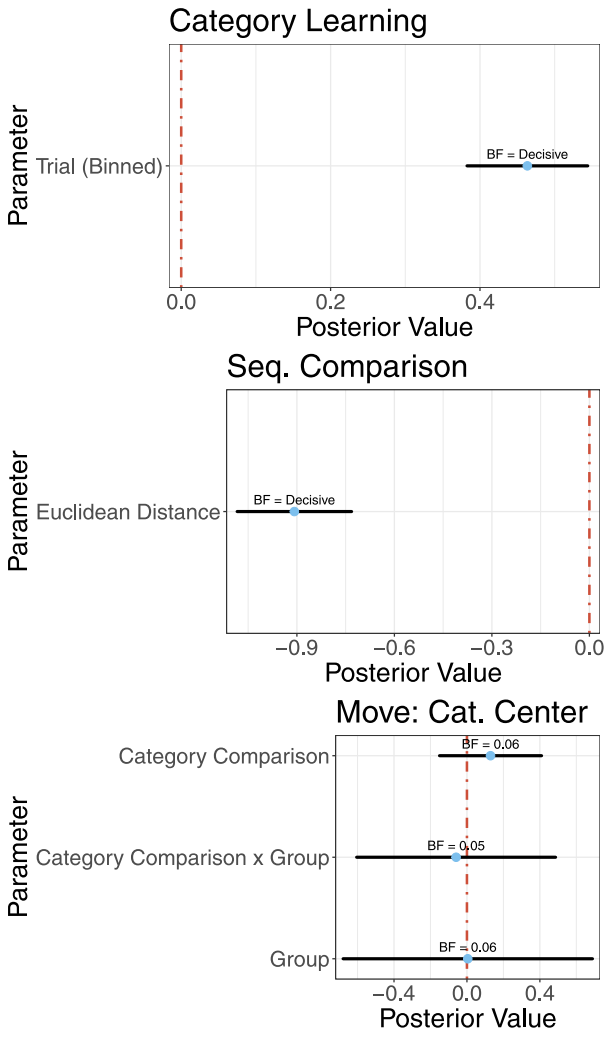


Fig. 25. Experiment 3.

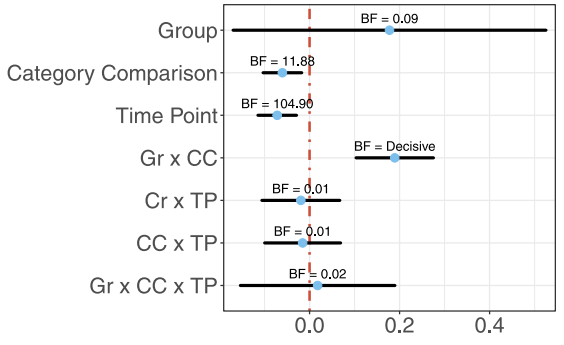


Fig. 26. Experiment 3 - c.

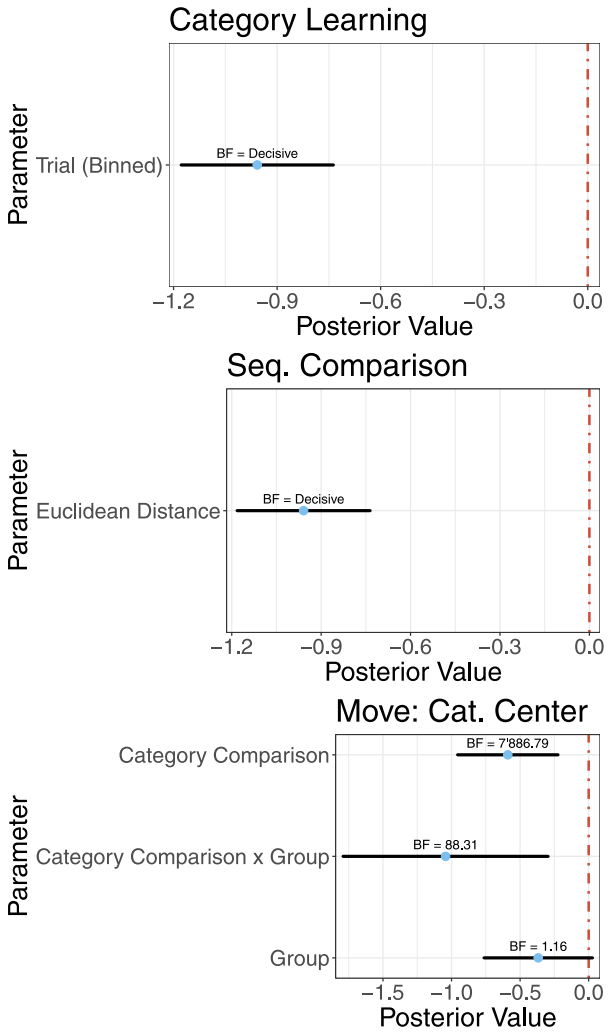


Fig. 27. Experiment 4.

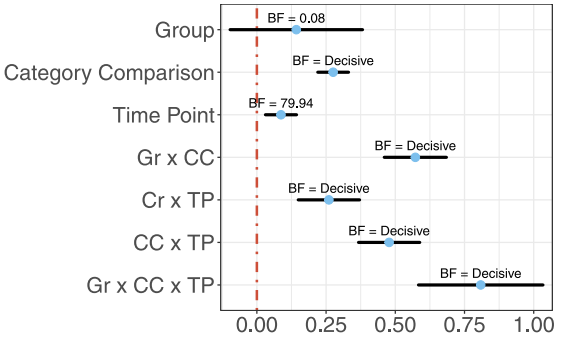


Fig. 28. Experiment 4 - c.

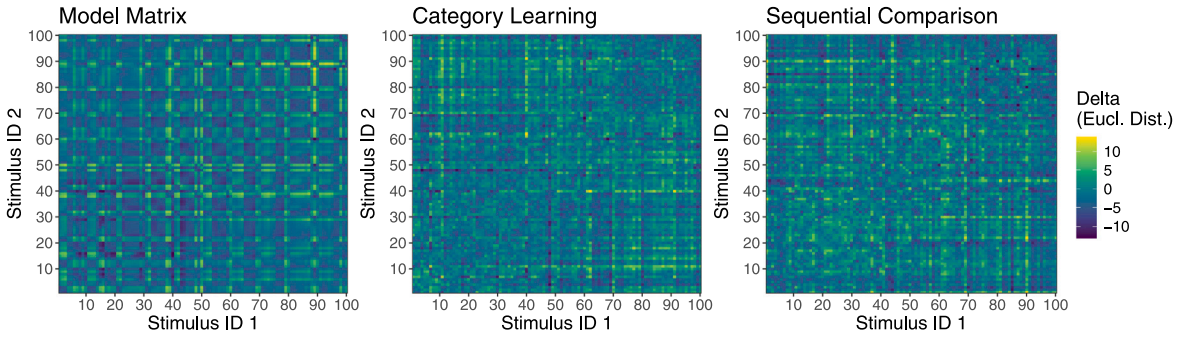


Fig. 29. Left: bRSA Model matrix of movements towards the corresponding category center from individual stimulus representations derived from model predictions for Experiment 1. Note. Lighter colors represent larger movements towards the corresponding category center, darker colors smaller movements. Middle: Average observed movements in the category learning group. Right: Average observed movements in the sequential comparison group.

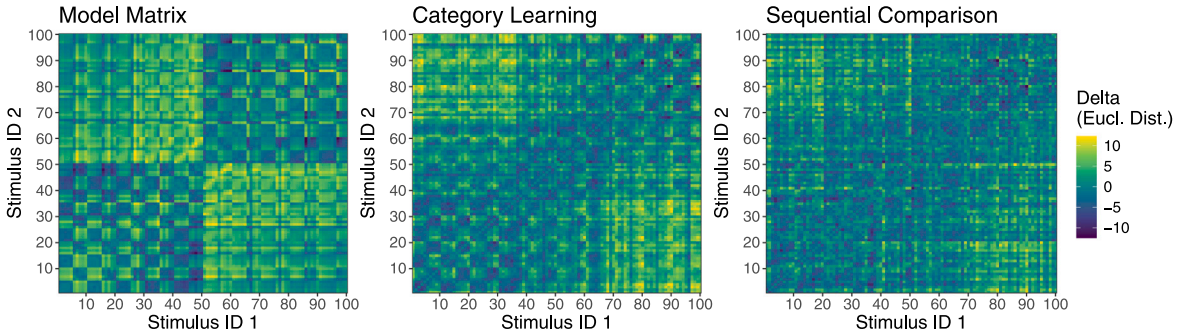


Fig. 30. Left: bRSA Model matrix of movements towards the corresponding category center from individual stimulus representations derived from model predictions for Experiment 2. Note. Lighter colors represent larger movements towards the corresponding category center, darker colors smaller movements. Middle: Average observed movements in the category learning group. Right: Average observed movements in the sequential comparison group.

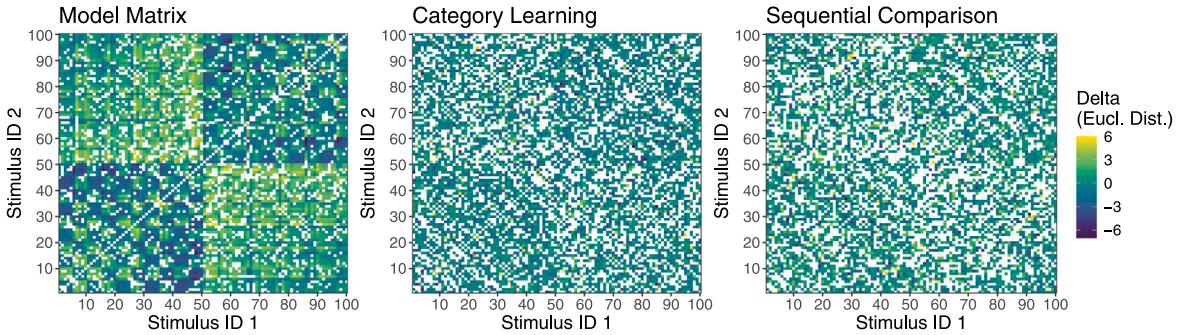


Fig. 31. Left: bRSA Model matrix of movements towards the corresponding category center from individual stimulus representations derived from model predictions for Experiment 3. Note. Lighter colors represent larger movements towards the corresponding category center, darker colors smaller movements. Middle: Average observed movements in the category learning group. Right: Average observed movements in the sequential comparison group.

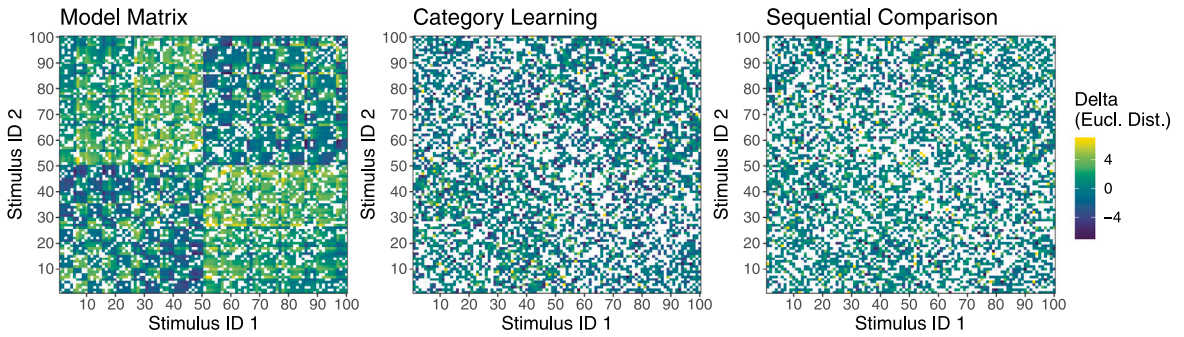


Fig. 32. Left: bRSA Model matrix of movements towards the corresponding category center from individual stimulus representations derived from model predictions for Experiment 4. Note. Lighter colors represent larger movements towards the corresponding category center, darker colors smaller movements. Middle: Average observed movements in the category learning group. Right: Average observed movements in the sequential comparison group.

References

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–485. http://dx.doi.org/10.1017/S0140525X00070801, Publisher: Cambridge University Press.

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. http://dx.doi.org/10.1037/0278-7393.14.1.33, Publisher: US: American Psychological Association.

Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of perceptual theory*: *Vol. 99*, *Advances in psychology* (pp. 369–399). North-Holland, http://dx.doi.org/10.1016/S0166-4115(08)62778-8.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154–179. http://dx.doi.org/10.1037/0033-295X.93.2.154, Place: US Publisher: American Psychological Association.

Barron, H. C., Garvert, M. M., & Behrens, T. E. J. (2016). Repetition suppression: a means to index neural representations using BOLD? *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *371*(1705), Article 20150355. http://dx.doi.org/10.1098/rstb.2015.0355.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*(1), 82–93. http://dx.doi.org/10.3758/BF03197629.

Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological Review*, *127*(5), 891–917. http://dx.doi.org/10.1037/rev0000197.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632–3645. http://dx.doi.org/10.1523/JNEUROSCI.3204-13.2014.

Bengio, Y., Courville, A., & Vincent, P. (2014). Representation learning: A review and new perspectives. http://dx.doi.org/10.48550/arXiv.1206.5538, URL http://arxiv.org/abs/1206.5538 arXiv:1206.5538 [cs].

Blaha, L. M., Busey, T. A., & Townsend, J. T. (2009). An LDA approach to the neural correlates of configural learning. *Vol. 31*, In *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 2540–2545).

Braida, L. D., & Durlach, N. I. (1972). Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, *51*(2B), 483–502. http://dx.doi.org/10.1121/1.1912868.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2022). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1).

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404. http://dx.doi.org/10.1016/0022-0965(82)90054-6.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. http://dx.doi.org/10.1016/0010-0285(73)90004-2, Number: 1.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. http://dx.doi.org/10.3758/s13428-014-0458-y.

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170–178. http://dx.doi.org/10.3758/s13423-014-0675-5.

Dubova, M., & Goldstone, R. L. (2021). The influences of category learning on perceptual reconstructions. *Cognitive Science*, *45*(5), http://dx.doi.org/10.1111/cogs.12981.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, *46*(2B), 372–383. http://dx.doi.org/10.1121/1.1911699.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39*, Article e229. http://dx.doi.org/10.1017/S0140525X15000965.

Gibson, E. J., & Walk, R. D. (1957). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology*, *49*(3), 239. http://dx.doi.org/10.1037/h0048274, Publisher: US: American Psychological Association.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 23.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43. http://dx.doi.org/10.1016/S0010-0277(00)00099-8.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116. http://dx.doi.org/10.1037/0096-3445.130.1.116, Publisher: US: American Psychological Association.

Harnad, S. (1995). In V. Honavar, & L. Uhr (Eds.), *Learned Categorical Perception in Neural Nets: Implications for Symbol Grounding* (pp. 191–206). Academic Press, Retrieved October 2, 2023, from https://eprints.soton.ac.uk/253357/.

Hasantash, M., & Afraz, A. (2020). Richer color vocabulary is associated with better color memory but not color perception. *Proceedings of the National Academy of Sciences*, *117*(49), 31046–31052. http://dx.doi.org/10.1073/pnas.2001946117.

Hebart, M. N., Zheng, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, *4*(11), 1173–1185. http://dx.doi.org/10.1038/s41562-020-00951-3, Number: 11 Publisher: Nature Publishing Group.

Holt, D. J., Boeke, E. A., Wolthusen, R. P. F., Nasr, S., Milad, M. R., & Tootell, R. B. H. (2014). A parametric study of fear generalization to faces and non-face objects: relationship to discrimination thresholds. *Frontiers in Human Neuroscience*, *8*, http://dx.doi.org/10.3389/fnhum.2014.00624, Publisher: Frontiers.

Homa, D., Sterling, S., & Trepel, L. (1982). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 418. http://dx.doi.org/10.1037/0278-7393.7.6.418, Publisher: US: American Psychological Association.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352. http://dx.doi.org/10.1037/0033-295X.98.3.352, Publisher: US: American Psychological Association.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220–241.

Johansen, M., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, *45*(4), 482–553. http://dx.doi.org/10.1016/S0010-0285(02)00505-4.

John, G. H., & Langley, P. (2013). Estimating continuous distributions in Bayesian classifiers. http://dx.doi.org/10.48550/arXiv.1302.4964, URL http://arxiv.org/abs/1302.4964 arXiv:1302.4964 [cs, stat].

Karagoz, A., Reagh, Z., & Kool, W. (2022). The construction and use of cognitive maps in model-based control. http://dx.doi.org/10.31234/osf.io/ngqwa.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572, Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1995.10476572.

Knoblauch, K., & Maloney, L. T. (2023). MLDS: Maximum likelihood difference scaling. Retrieved April 8, 2024, from https://cran.r-project.org/web/packages/MLDS/index.html.

Kruschke, J. K. (2005). Learning involves attention. *Connectionist Models in Cognitive Psychology*, *51*, 113–140.

Lewandowsky, S. (1999). Redintegration and response suppression in serial recall: A dynamic network model. *International Journal of Psychology*, *34*(5–6), 434–446. http://dx.doi.org/10.1080/002075999399792, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1080/002075999399792.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. http://dx.doi.org/10.1037/h0044417.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*(3), 1242–1255. http://dx.doi.org/10.1121/1.408177.

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, *6*(3), 312–319. http://dx.doi.org/10.3758/BF03197461, Number: 3.

Logan, G. D. (1989). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492. http://dx.doi.org/10.1037/0033-295X.95.4.492, Publisher: US: American Psychological Association.

Luce, R., Green, D., & Weber, D. (1976). Attention bands in absolute identification. *Perception & Psychophysics*, *20*(1), 49–54. http://dx.doi.org/10.3758/BF03198705.

Luce, R., Nosofsky, R., Green, D., & Smith, A. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics*, *32*(5), 397–408. http://dx.doi.org/10.3758/BF03202769.

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. http://dx.doi.org/10.1073/pnas.1614048113, Publisher: Proceedings of the National Academy of Sciences.

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. http://dx.doi.org/10.1016/j.neulet.2017.07.061, URL https://www.sciencedirect.com/science/article/pii/S030439401730647X.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 227–245.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision*, *3*(8), 5. http://dx.doi.org/10.1167/3.8.5.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. http://dx.doi.org/10.1016/j.jml.2017.01.001.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*(2), 254–278. http://dx.doi.org/10.1037/0033-295X.100.2.254, Place: US Publisher: American Psychological Association.

Milton, F., & Pothos, E. M. (2011). Category structure and the two learning systems of COVIS. *European Journal of Neuroscience*, *34*(8), 1326–1336. http://dx.doi.org/10.1111/j.1460-9568.2011.07847.x, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-9568.2011.07847.x.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775. http://dx.doi.org/10.1037/0278-7393.27.3.775, Publisher: US: American Psychological Association.

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2024). Bayesfactor: Computation of Bayes factors for common designs. Retrieved April 5, 2024, from https://cran.r-project.org/web/packages/BayesFactor/index.html.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship.. *Journal of Experimental Psychology: General*, *115*(1), 39. http://dx.doi.org/10.1037/0096-3445.115.1.39, Publisher: US: American Psychological Association.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 3–27.

Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition*, *33*(7), 1256–1271. http://dx.doi.org/10.3758/BF03193227.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 924. http://dx.doi.org/10.1037/0278-7393.28.5.924, Publisher: US: American Psychological Association.

Owen, D. H., & Machamer, P. K. (1979). Bias-free improvement in wine discrimination. *Perception*, *8*(2), 199–209. http://dx.doi.org/10.1068/p080199, Publisher: SAGE Publications Ltd STM.

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1224–1231. http://dx.doi.org/10.1037/a0030947.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353–363. http://dx.doi.org/10.1037/h0025953.

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, *1*(2), 125–132. http://dx.doi.org/10.1038/35039062, Bandiera_abtest: a Cg_type: Nature Research Journals Number: 2 Primary_atype: Reviews Publisher: Nature Publishing Group.

R Core Team (2022). *R: A language and environment for statistical computing*. URL https://www.R-project.org/.

Rahnev, D., Block, N., Denison, R. N., & Jehee, J. (2021). Is perception probabilistic? Clarifying the definitions. http://dx.doi.org/10.31234/osf.io/f8v5r.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*, Article e223. http://dx.doi.org/10.1017/S0140525X18000936.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *82*(2).

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, *8*(1), 118–137. http://dx.doi.org/10.1111/tops.12174.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). MIT Press, Google-Books-ID: k5Sr0nFw7psC.

Sims, C. R. (2016). Rate–distortion theory and human perception. *Cognition*, *152*, 181–198. http://dx.doi.org/10.1016/j.cognition.2016.03.020.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1411. http://dx.doi.org/10.1037/0278-7393.24.6.1411, Publisher: US: American Psychological Association.

Souza, A. S., Overkott, C., & Matyja, M. (2021). Categorical distinctiveness constrains the labeling benefit in visual working memory. *Journal of Memory and Language*, *119*, Article 104242. http://dx.doi.org/10.1016/j.jml.2021.104242.

Souza, A. S., Rerko, L., Lin, H.-Y., & Oberauer, K. (2014). Focused attention improves working memory: implications for flexible-resource and discrete-capacity models. *Attention, Perception, & Psychophysics*, *76*(7), 2080–2102. http://dx.doi.org/10.3758/s13414-014-0687-2.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition*, *166*, 277–297. http://dx.doi.org/10.1016/j.cognition.2017.05.038, Place: Netherlands Publisher: Elsevier Science.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37–55. http://dx.doi.org/10.1037/xlm0000578, Place: US Publisher: American Psychological Association.

Theves, S., Fernandez, G., & Doeller, C. F. (2019). The Hippocampus encodes distances in multidimensional feature space. *Current Biology*, *29*(7), 1226–1231.e3. http://dx.doi.org/10.1016/j.cub.2019.02.035.

Theves, S., Fernández, G., & Doeller, C. F. (2020). The Hippocampus maps concept space, not feature space. *The Journal of Neuroscience*, *40*(38), 7318–7325. http://dx.doi.org/10.1523/JNEUROSCI.0494-20.2020, URL https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0494-20.2020.

Theves, S., Neville, D. A., Fernández, G., & Doeller, C. F. (2021). Learning and representation of hierarchical concepts in hippocampus and prefrontal cortex. *The Journal of Neuroscience*, *41*(36), 7675–7686. http://dx.doi.org/10.1523/JNEUROSCI.0657-21.2021.

Vanpaemel, W., & Navarro, D. J. (2007). Representational shifts during category learning. In *Proceedings of the 29th annual conference of the cognitive science society* (pp. 1599–1604).

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. http://dx.doi.org/10.1007/s11222-016-9696-4.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535–585. http://dx.doi.org/10.1016/0010-0285(92)90018-W.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. http://dx.doi.org/10.3758/BF03194105, Number: 5.

Wetzels, R., Raaijmakers, J. G. W., Jakab, E., & Wagenmakers, E.-J. (2009). How to quantify support for and against the null hypothesis: A flexible Winbugs implementation of a default Bayesian t test. *Psychonomic Bulletin & Review*, *16*(4), 752–760. http://dx.doi.org/10.3758/PBR.16.4.752.