# Category-Specific Sensitivity Allows Modeling Variability Effects in a Similarity Framework

**Florian I. Seitz (florian.seitz@unibas.ch)**
Department of Psychology, Missionsstrasse 62A
Basel, Switzerland

**Rebecca Albrecht (rebecca.albrecht@cognition.uni-freiburg.de)**
Center for Cognitive Science, Hebelstraße 10
Freiburg, Germany

## Abstract

Categorization is a core cognitive ability that enables humans to treat distinct entities as equivalent members of the same concept. Similarity-based models have been highly successful in explaining a wide range of categorization phenomena, but struggle to account for the category variability effect—the tendency to assign ambiguous items to more variable categories—because they do not explicitly represent distributional properties such as variability. We argue that explaining this effect does not require abandoning similarity-based models, but rather relaxing the assumption that model parameters are fixed across categories. Specifically, we consider two parametric extensions: allowing a bias parameter to favor high-variability categories, and allowing a sensitivity parameter to vary with each category's distributional precision. We directly compare both mechanisms in a model simulation across prototype and exemplar models, evaluating their categorization performance. Results consistently favor category-specific sensitivity over response bias, particularly for prototype models. We interpret this finding in terms of the broader cognitive science literature and extend the sensitivity account to prevalence-induced concept change, arguing that both variability and prevalence effects may reflect a common underlying mechanism in which sensitivity scales with the degree to which a category is specified. Together, these findings point towards a unified similarity-based account of distributional influences on categorization.

**Keywords:** Similarity; Perceptual categorization; Computational modeling; Category variability effect; Prevalence-induced concept change

## Introduction

Categorization is a fundamental component of cognition that enables people to treat distinct entities as equivalent members of the same concept. Real-world categories differ in distributional properties like variability and prevalence, and people are sensitive to such information (e.g., Maddox, 1995; E. E. Smith & Sloman, 1994). For example, there is broad consensus that dogs vary more than cats in size, weight, and other features, and are encountered more frequently than many other species. Similarly, political views are distributed differently across groups, with moderate groups encompassing a broader range of positions and occurring more frequently than ideologically extreme groups. Understanding perceptual and social concepts therefore requires understanding how the mind incorporates distributional information during categorization. The present work examines the role of variability within a leading theoretical framework that explains categorization in terms of similarity.

The similarity framework assumes an item is most likely assigned to the most similar category (Medin & Schaffer, 1978). Categories, in turn, are typically represented either by their central tendency (the *prototype*; D. J. Smith & Minda, 1998) or by the individual category members (the *exemplars*; Nosofsky, 1986). By computing similarity to prototypes or exemplars, similarity-based models have been highly successful in accounting for various categorization phenomena (Goldstone, 1994), including typicality effects and graded category structure (Nosofsky, 1988; Rosch & Mervis, 1975), generalization to novel items (Nosofsky, 1986; Shepard, 1987), knowledge abstraction (Posner & Keele, 1968; D. J. Smith & Minda, 1998), and the handling of exceptions (Nosofsky, Palmeri, & McKinley, 1994; Nosofsky & Palmeri, 1998).

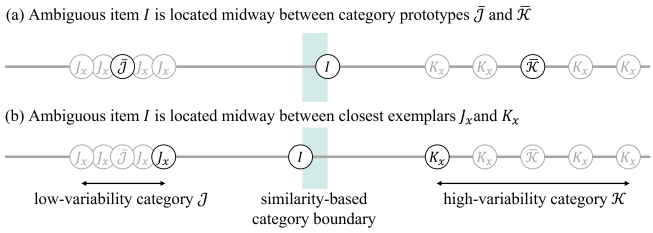


Figure 1: *Category variability effect*. When two categories differ in variability, people tend to assign an ambiguous item $I$ located midway between prototypes (a) or nearest exemplars (b) to the high-variability category $\mathcal{K}$. In (b), this preference conflicts with the possible space of category boundaries predicted by standard similarity-based models (green ribbon).

Despite these successes, similarity-based theories consider distributional information such as category variability to a surprisingly limited extent (cf. E. E. Smith & Sloman, 1994): Prototype theory disregards such information altogether, whereas exemplar theory incorporates it only implicitly through the accumulation of stored exemplars. This limitation is brought into sharp focus by the *category variability effect*, in which ambiguous items are preferentially assigned to more variable categories (see Fig. 1; e.g., Cohen, Nosofsky, & Zaki, 2001)—a pattern that standard similarity-based models fail to predict, or even predict in the wrong direction (cf. Fig. 1b).

Prior work has proposed two parametric mechanisms through which the similarity framework could be extended to accommodate this effect: a bias favoring high-variability



In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* 

categories (Cohen et al., 2001) and increased sensitivity to differences from low-variability categories (Nosofsky & Johansen, 2000). However, these accounts have not been directly compared. The present work addresses this gap by fitting both mechanisms within prototype and exemplar models and evaluating their ability to support accurate categorization. This approach allows us to assess which parameterization provides the more robust account of the category variability effect and to examine how the parameters adapt when optimized for categorization accuracy. Our results consistently favor category-specific sensitivity over response bias, particularly within prototype models. We then extend the sensitivity account to a related but distinct phenomenon—*prevalence-induced concept change* (Levari et al., 2018)—arguing that a unified sensitivity-based mechanism may explain distributional influences on categorization more broadly.

## Similarity Framework: Core Assumptions

Similarity-based models of categorization assume that items that are more similar are more likely to belong to the same category (D. J. Smith & Minda, 1998; Nosofsky, 1986). Accordingly, the probability $\Pr(\mathcal{J}|I)$ that an item $I$ is assigned to a category $\mathcal{J}$ increases with the corresponding similarity $s_{I\mathcal{J}}$. Following Luce's (1959) choice rule, $\Pr(\mathcal{J}|I)$ is given by

$$\Pr(\mathcal{J}|I) = \frac{s_{I\mathcal{J}}}{\sum_{\mathcal{K}} s_{I\mathcal{K}}}, \tag{1}$$

where $\mathcal{K}$ refers to all considered category alternatives. Similarity $s_{I\mathcal{J}}$ is typically defined in one of two ways (Nosofsky, 1992; Nosofsky & Zaki, 2002): In exemplar models, $s_{I\mathcal{J}}$ reflects the summed similarity between item $I$ and all $n_{\mathcal{J}}$ exemplars $J_x$ associated with category $\mathcal{J}$ ($s_{I\mathcal{J}} = \sum_{J_x \in \mathcal{J}} s_{IJ_x}$). Prototype models, in turn, use the similarity between $I$ and the category's prototype $\bar{\mathcal{J}}$ ($s_{I\mathcal{J}} = s_{I\bar{\mathcal{J}}}$, with $\bar{\mathcal{J}} = \frac{1}{n_{\mathcal{J}}} \sum_{J_x \in \mathcal{J}} J_x$).

Grounded in Shepard (1987), similarity between items is commonly modeled as a negative exponential function of their feature value differences (Nosofsky, 1986). In the present work, we focus on items defined along a single, continuous feature, such that the difference $d_{IJ}$ between items $I$ and $J$ is given by $d_{IJ} = |I - J|$. Accordingly, the similarity $s_{IJ}$ between $I$ and $J$—where $J$ may denote a category exemplar $J_x$ or prototype $\bar{\mathcal{J}}$—is defined as

$$s_{IJ} = \exp(-c \cdot d_{IJ}), \tag{2}$$

where $c > 0$ is a free sensitivity parameter that determines how steeply similarity declines with increasing feature value differences (see Fig. 2a). Larger values of $c$ produce a steeper similarity gradient, such that identical physical differences can correspond to markedly smaller psychological similarities with increasing $c$.

As can be seen from Eqs. 1 and 2, the similarity framework predicts responses without explicitly representing category variability. Prototype models neglect any distributional information beyond central tendencies and predict stable linear category boundaries located midway between prototypes (in Fig. 1, the predicted category boundary is located at the right end of the green ribbon). Consequently, each new item is assigned to the category with the closest prototype, irrespective of differences between categories in variability or other distributional properties. Exemplar models, in turn, incorporate distributional information only implicitly by summing similarity across exemplars. As a result, they favor high-prevalence categories (in general compatible with human categorization behavior; e.g., Maddox, 1995; D.-Y. Yang & Worthy, 2026) and predict different effects for variability depending on how ambiguous items are placed between categories (cf. Fig. 1). Specifically, exemplar models predict that ambiguous items are assigned to a high-variability category when prototypes are held at a constant distance from them (Fig. 1a), but to a low-variability category when the closest exemplars are held at a constant distance (Fig. 1b).

## The Category Variability Effect: An Empirical Challenge for the Similarity Framework

Despite its success in explaining categorization (e.g., Nosofsky, 1986), the similarity framework has been challenged by empirical findings related to category variability, which this work aims to reconcile with a similarity-based account.

In studies on category variability, two categories differ in variability and, usually, test items of interest are located midway between the closest exemplars of the two categories (see Fig. 1b). Although such an ambiguous item is closer and more similar to the low-variability category (both to exemplars and prototypes), people often assign it to the high-variability category. This *category variability effect* is therefore incompatible with standard similarity-based predictions (E. E. Smith & Sloman, 1994), and has been reported in multiple studies, albeit with considerable variation in magnitude. For example, Perlman, Hahn, Edwards, and Pothos (2012) and L.-X. Yang and Huang (2021) found that roughly 75% of participants showed the effect, which, however, disappeared when variability differences between categories were made less salient (Stewart & Chater, 2002). The effect also appears to vanish for stimuli with multiple features, possibly due to the more complex categorization task (Cohen et al., 2001; Seitz, Jarecki, & Rieskamp, 2023). Moreover, L.-X. Yang and Wu (2014) and Seitz, Jarecki, and Rieskamp (2023) reported substantial interindividual differences, suggesting that the category variability effect is robust but strongly modulated by task structure and individual strategies.

### Accounts for the Category Variability Effect

This section describes two mechanisms through which similarity-based models can predict choosing the high-variability category for ambiguous test items (also discussed in Nosofsky & Johansen, 2000; Cohen et al., 2001). The goal is to evaluate which provides the more robust account of the effect—a question we examine through model fitting and performance comparisons in a simulation detailed below.

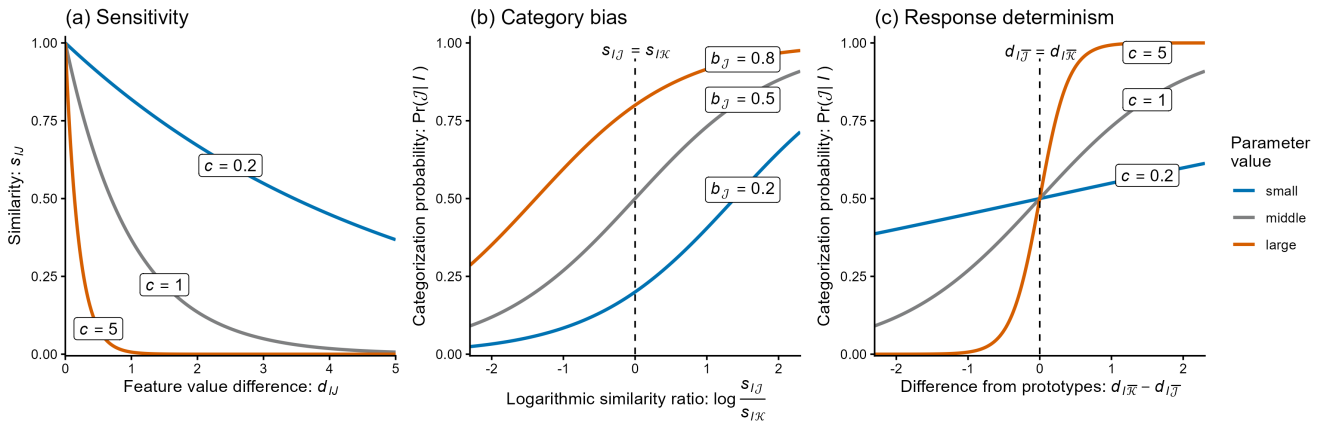


Figure 2: *Parameters of the similarity framework in a binary categorization task.* Panel (a) shows how a larger sensitivity $c$ makes similarity decrease more steeply with feature value difference (Eq. 2). Panels (b) and (c) show the effect of parameters on the probability for selecting category $\mathcal{J}$ which, according to Eq. 1, depends on the similarity ratio $s_{I\mathcal{J}}/s_{I\mathcal{K}}$. In panel (b), a larger bias $b_{\mathcal{J}}$ shifts categorization towards category $\mathcal{J}$. Panel (c) shows that, in a prototype model, sensitivity $c$ can alternatively be understood as temperature of a softmax, with larger values making categorization more deterministic.

**Sensitivity to Low-Variability Categories.** One way to favor high-variability categories is through increased *sensitivity to low-variability categories* (Nosofsky & Johansen, 2000). This greater sensitivity implies that narrow categories make items with differing feature values appear particularly dissimilar (see Fig. 2a). Formally, allowing the sensitivity parameter $c$ to vary across categories extends Eq. 2 to

$$s_{IJ} = \exp\left(-c_{\mathcal{J}} \cdot d_{IJ}\right), \tag{3}$$

where $c_{\mathcal{J}}$ denotes the sensitivity to differences from item $J$ associated with category $\mathcal{J}$. An ambiguous item is more similar—and therefore more likely assigned—to a high-variability category $\mathcal{H}$ than to a low-variability category $\mathcal{L}$ when $c_{\mathcal{L}}$ is sufficiently larger than $c_{\mathcal{H}}$. Psychologically, this corresponds to people being highly sensitive to differences when evaluating narrow categories, alongside a relative discounting of differences for highly variable categories.

**Bias for High-Variability Categories.** An alternative way to predict the category variability effect is through a *bias for high-variability categories* (Cohen et al., 2001). In similarity-based models, category biases modulate the extent to which similarity evidence can override default category responses. Formally, introducing category biases extends Eq. 1 to

$$\Pr(\mathcal{J}|I) = \frac{b_{\mathcal{J}} \cdot s_{I\mathcal{J}}}{\sum_{\mathcal{K}} b_{\mathcal{K}} \cdot s_{I\mathcal{K}}}, \tag{4}$$

where $b_{\mathcal{J}}$ denotes the bias associated with category $\mathcal{J}$ (with $0 \leq b_{\mathcal{J}} \leq 1$ for all $b_{\mathcal{J}}$ and $\sum_{\mathcal{J}} b_{\mathcal{J}} = 1$). An ambiguous item is likely assigned to a high-variability category $\mathcal{H}$, if $\mathcal{H}$ is associated with a large bias $b_{\mathcal{H}}$ (see Fig. 2b). Psychologically, this corresponds to a general tendency to favor broad, variable categories unless the similarity evidence strongly supports assignment to a narrow, highly specified category.

**Relationship to prior work.** Both accounts have individual precedents: Nosofsky and Johansen (2000) introduced the sensitivity account in an appendix simulation, and Cohen et al. (2001) developed the bias account and argued it could improve overall categorization accuracy. However, neither study directly compared the two mechanisms against each other across both prototype and exemplar models. The present simulations address this gap by evaluating which mechanism provides better categorization performance across varying degrees of category variability.

It could be tempting to interpret category-specific sensitivity as reflecting adaptations of the cognitive representation itself, whereas category biases might instead reflect relatively simple adjustments of the choice rule (cf. Rosch, 1978). However, this distinction is difficult to maintain due to substantial overlap between representational and decisional components of similarity-based models. In a prototype model, Eqs. 1 and 2 can be summarized to a softmax function

$$\Pr(\mathcal{J}|I) = \frac{\exp\left(-c \cdot d_{I\bar{\mathcal{J}}}\right)}{\sum_{\mathcal{K}} \exp\left(-c \cdot d_{I\bar{\mathcal{K}}}\right)} \tag{5}$$

$$= \frac{1}{1 + \sum_{\mathcal{K} \neq \mathcal{J}} \exp\left(-c \cdot \left[d_{I\bar{\mathcal{K}}} - d_{I\bar{\mathcal{J}}}\right]\right)}, \tag{6}$$

where the sensitivity parameter $c$ corresponds to the temperature of softmax. Rather than being intrinsic to cognitive processes, parameter $c$ can thus also be interpreted as governing response determinism (see Fig. 2c). From this perspective, explaining a preference for high-variability categories in terms of sensitivity can be reinterpreted as differential response determinism: Similarity evidence is used more deterministically when it favors more variable categories, perhaps because these categories are perceived as more likely to accommodate additional items.

Table 1: Optimal estimates for bias $b$ and sensitivity $c$, given a low-variability category $\mathcal{L}$ with $\sigma_{\mathcal{L}} \leq \sigma_{\mathcal{H}} = 1$ of category $\mathcal{H}$.

| | Prototype model | | | | | | Exemplar model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_{\mathcal{L}}$ | $b_{\mathcal{L}}$ | $b_{\mathcal{H}}$ | $c_{\mathcal{L}}$ | $c_{\mathcal{H}}$ | $\ell$ | % corr. | $b_{\mathcal{L}}$ | $b_{\mathcal{H}}$ | $c_{\mathcal{L}}$ | $c_{\mathcal{H}}$ | $\ell$ | % corr. |
| *baseline model: no category-specific parameters* | | | | | | | | | | | | |
| 0.01 | 0.50 | 0.50 | 2.05 | 2.05 | -78.65 | 84.47 | 0.50 | 0.50 | 20.00 | 20.00 | -26.22 | 95.06 |
| 0.10 | 0.50 | 0.50 | 2.08 | 2.08 | -80.82 | 84.47 | 0.50 | 0.50 | 19.56 | 19.56 | -42.20 | 92.61 |
| 1.00 | 0.50 | 0.50 | 1.04 | 1.04 | -119.98 | 68.83 | 0.50 | 0.50 | 3.15 | 3.15 | -117.57 | 68.62 |
| *category-specific bias b* | | | | | | | | | | | | |
| 0.01 | 0.00 | 1.00 | 14.50 | 14.50 | -50.43 | 91.27 | 0.28 | 0.72 | 20.00 | 20.00 | -21.97 | 96.37 |
| 0.10 | 0.08 | 0.92 | 4.41 | 4.41 | -63.84 | 88.43 | 0.44 | 0.56 | 19.26 | 19.26 | -40.95 | 92.85 |
| 1.00 | 0.50 | 0.50 | 1.04 | 1.04 | -119.89 | 68.91 | 0.50 | 0.50 | 3.19 | 3.19 | -117.42 | 68.64 |
| *category-specific sensitivity c* | | | | | | | | | | | | |
| 0.01 | 0.50 | 0.50 | 20.00 | 2.78 | -20.58 | 96.62 | 0.50 | 0.50 | 20.00 | 8.00 | -21.29 | 96.53 |
| 0.10 | 0.50 | 0.50 | 15.36 | 3.58 | -40.06 | 92.79 | 0.50 | 0.50 | 19.66 | 14.48 | -40.34 | 92.94 |
| 1.00 | 0.50 | 0.50 | 1.04 | 1.04 | -119.86 | 68.86 | 0.50 | 0.50 | 3.20 | 3.23 | -117.42 | 68.61 |
| *category-specific bias b and sensitivity c* | | | | | | | | | | | | |
| 0.01 | 0.62 | 0.38 | 20.00 | 2.15 | -20.32 | 96.67 | 0.64 | 0.36 | 20.00 | 7.31 | -20.55 | 96.67 |
| 0.10 | 0.89 | 0.11 | 16.19 | 1.18 | -38.50 | 93.52 | 0.80 | 0.20 | 19.92 | 5.01 | -39.10 | 93.10 |
| 1.00 | 0.50 | 0.50 | 1.05 | 1.04 | -119.49 | 68.89 | 0.50 | 0.50 | 3.44 | 3.66 | -116.95 | 68.81 |

*Note.* $\ell$, log likelihood of the true category given the model predictions; % corr, percentage of correctly predicted modal categorizations.

## Simulating Category-Specific Sensitivity and Bias

The accounts described above have been proposed on theoretical grounds, but an important open question is which mechanism—sensitivity or bias—allows the similarity framework to more accurately classify items from categories that differ in variability. We address this by examining how category variability shapes optimal parameter values when classifying new items. The code for the simulation reported below can be found in the Github files associated with https://osf.io/rzdw6/overview.

**Simulation setup.** The simulation considers objects with a single feature belonging to either a low-variability category $\mathcal{L}$ or a high-variability category $\mathcal{H}$. Given the category prototype or 100 exemplars randomly drawn from each category, the parameter values that maximize the likelihood of correctly classifying 100 new items drawn from each category were estimated. The model contained two free parameters (sensitivity $0 \leq c \leq 20$ and bias $0 \leq b \leq 1$) and four paramtrization variants: (a) a baseline where parameters did not differ across categories, (b) a variant with unequal category biases, (c) a variant with category-specific sensitivity, and (d) a variant with both category-specific biases and sensitivity. Based on Nosofsky and Johansen (2000), the high-variability category $\mathcal{H}$ was modeled as a normal distribution with a mean feature value of 1 (i.e., the prototype) and a standard deviation of 1; the low-variability category $\mathcal{L}$ as a normal distribution with a mean feature value of 0 (i.e., the prototype) and a standard deviation $\sigma_{\mathcal{L}}$ of .01, .1, or 1. These three values were chosen to span a wide range of variability contrasts: $\sigma_{\mathcal{L}} = .01$ represents an extremely narrow category with a 100:1 variability ratio, $\sigma_{\mathcal{L}} = .1$ represents a moderately narrow category with a 10:1 variability ratio, and $\sigma_{\mathcal{L}} = 1$ represents a control condition with equal variability in both categories. Examining all three conditions together therefore reveals whether the mechanisms accommodate variability differences in a graded, principled way. For each parametrization variant and standard deviation $\sigma_{\mathcal{L}}$, the simulation was repeated 100 times; each row in Table 1 reports the mean parameter estimates and performance metrics across repetitions for that combination, separately for prototype and exemplar models.

**Results.** Relative to the baseline model with category-invariant parameters, allowing category-specific parameters substantially improved model performance—but primarily for prototype models. In prototype models, introducing category-specific sensitivity led to marked gains in accuracy, improving classification by more than 10 percentage points and outperforming models with category-specific response biases. Sensitivity-based prototype models matched the performance of all exemplar-model variants, whereas category biases alone yielded comparatively modest improvements. By contrast, exemplar models benefited little from category-specific parameters, consistent with the fact that they implicitly encode variability information through the accumulation of stored exemplars. Allowing both sensitivity and bias to vary across categories did not further improve model fit beyond sensitivity alone; moreover, the direction of the estimated bias reversed, indicating instability in bias estimation when both parameters are allowed to vary across categories. When categories were equally vari-

able ($\sigma_{\mathcal{L}} = 1$), optimal parameters converged to the symmetric baseline in all conditions—confirming that the advantage of category-specific sensitivity is specific to variability differences, not an artifact of added model flexibility. Taken together, these results suggest that category-specific sensitivity provides a more robust and principled mechanism than response biases for capturing variability effects, particularly within prototype-based accounts.

## Discussion

Similarity provides a powerful framework for modeling categorization, yet it has long struggled to explain how people incorporate distributional information such as category variability into their decisions (E. E. Smith & Sloman, 1994). This article reviewed how the similarity framework can nevertheless account for the category variability effect, whereby ambiguous items located midway between two category edges are preferentially assigned to the more variable category. We showed that this effect can be explained by a response bias favoring the high-variability category or by increased sensitivity to differences from the low-variability category (Cohen et al., 2001; Nosofsky & Johansen, 2000). Our simulations, which compared these two mechanisms in terms of their categorization performance, indicate that increased sensitivity provides a higher-performing and methodologically more robust account, particularly for prototype models.

This finding aligns with prior work suggesting that sensitivity may be flexibly adjusted to task demands (Lamberts, 1994; Seitz, Jarecki, Rieskamp, & von Helversen, 2025). For example, sensitivity tends to be higher in tasks requiring fine-grained discrimination, such as old–new recognition, than in broader categorization tasks that emphasize grouping rather than differentiation (e.g., Shin & Nosofsky, 1992). Moreover, effects previously attributed to response determinism (e.g., Seitz, von Helversen, Albrecht, Rieskamp, & Jarecki, 2023) might instead be explained in terms of sensitivity (see Eq. 6; Krefeld-Schwalb, Pachur, & Scheibehenne, 2022).

### Relations to Bayesian and Connectionist Accounts

The sensitivity account is conceptually related to approaches that treat categorization as inference over generative models of category distributions. In Bayesian accounts, the probability of assigning item $I$ to category $\mathcal{J}$ is proportional to the likelihood $p(I|\mathcal{J})$ times the prior $p(\mathcal{J})$ (Anderson, 1991; Tenenbaum & Griffiths, 2001). When categories are Gaussian, $p(I|\mathcal{J})$ depends on the distance from the category mean scaled by variance—effectively mirroring Eq. 3. The sensitivity account proposed here can thus be understood as a similarity-based approximation to Bayesian likelihood weighting: higher sensitivity to a low-variance category corresponds to the steeper likelihood gradient that Bayesian inference naturally assigns to such categories. This connection suggests that the sensitivity account is not merely a post-hoc correction to the similarity framework, but is in fact what a principled probabilistic reasoner would do when categories differ in variability.

The relation to connectionist models is equally informative. In radial basis function networks—a class of neural networks closely related to exemplar models (Kruschke, 1992; Jäkel, Schölkopf, & Wichmann, 2008)—each hidden unit computes a Gaussian-shaped similarity to a stored exemplar, and the width of this Gaussian (the inverse of sensitivity) can be learned from experience. When trained on categories that differ in variability, the resulting receptive fields are narrower for low-variability categories and broader for high-variability categories (Poggio & Girosi, 1990), mirroring differential sensitivity in similarity-based models.

### Additional Similarity-Based Explanations

Beyond the parametric accounts of the category variability effect examined above, non-parametric explanations grounded in the similarity framework have also been proposed. We briefly review these to situate our contribution within the broader literature, while noting that a systematic comparison of parametric and non-parametric accounts remains an important direction for future work.

One intuitive account relies on standardized similarity measures, such as the Mahalanobis similarity, which scale feature value differences by category-specific variability as in likelihood gradients in Bayesian inference (Battleday, Peterson, & Griffiths, 2020; Seitz, Jarecki, & Rieskamp, 2026). For single-feature items, similarity depends on the standardized difference $d_{IJ} = |I - J|/\sigma_{\mathcal{J}}$, where $\sigma_{\mathcal{J}}$ denotes the feature's standard deviation within category $\mathcal{J}$. This formulation directly incorporates variability: although an ambiguous item midway between category boundaries is closer in absolute terms to the low-variability category, it lies fewer standard deviations from the high-variability category, yielding the category variability effect. Standardized similarity also accounts for more subtle distributional effects, including illusory variability induced by presentation order (Sakamoto, Jones, & Love, 2008) and sensitivity to feature correlations in multidimensional spaces (Seitz et al., 2026). Notably, the Mahalanobis approach can be understood as a non-parametric implementation of category-specific sensitivity in which sensitivity is determined entirely by observed category statistics, without requiring additional free parameters.

Another approach focuses on sequence effects—the influence of preceding items on current categorizations (Stewart & Chater, 2002; Stewart & Brown, 2004). Stewart, Brown, and Chater (2002) proposed the memory-and-comparison strategy, a sequential heuristic in which the current item, again consisting of a single feature, is compared to the immediately preceding one. If the direction of change unambiguously signals category membership—for instance, when the feature value of the previous item exceeded the category boundary and the current item's feature value is even larger—categorization is deterministic. When this heuristic does not apply, the model assumes the previous category label is repeated with a probability equal to the similarity between the successive items. As a consequence, ambiguous items may be classified differently depending on the preceding item, poten-

tially canceling out the category variability effect when averaged across trials—a pattern in line with the results of several studies (Cohen et al., 2001; Perlman et al., 2012; L.-X. Yang & Wu, 2014). Consistent with this prediction, L.-X. Yang and Wu (2014) identified a subgroup of participants whose category variability effect depended systematically on stimulus sequence rather than reflecting a stable preference.

Finally, a related line of research on exemplar retrieval suggests that behavior often relies on only a few highly similar exemplars (Albrecht, Hoffmann, Pleskac, Rieskamp, & von Helversen, 2020; Nosofsky & Palmeri, 1997; Seitz, Albrecht, von Helversen, Rieskamp, & Rosner, 2025). Rather than aggregating similarity across all exemplars, a competitive retrieval process selects a single exemplar per category with retrieval probability proportional to similarity. Formally, the probability of retrieving exemplar $J_x$ when categorizing item $I$ is given by $\frac{s_{IJ_x}}{\sum_{J_x \in \mathcal{J}} s_{IJ_x}}$, where the denominator normalizes the probabilities to sum to 1 across the exemplars of a category. Such competitive retrieval can produce the category variability effect: When the ambiguous item is located midway between the nearest exemplars, the nearest exemplar from the high-variability category has a higher normalized retrieval probability than the nearest exemplar from the low-variability category due to the different normalization terms. By relying on a single retrieved exemplar per category, this approach occupies a conceptual middle ground between exemplar and prototype models.

Taken together, these non-parametric accounts show that the category variability effect can emerge from how similarity is computed, contextualized, or used for retrieval. What they share with the parametric accounts examined here is a common intuition: ambiguous items should be judged relative to the distributional structure of each category.

## Application to Prevalence Effects

Variability effects are closely related to prevalence effects, as both concern how distributional information shapes categorization. Whereas variability determines how broadly a category spans feature space, prevalence determines how much evidence is available for that category. Both properties reflect the degree to which a category is specified: a low-variability category is tightly specified in feature space, and a high-prevalence category is richly specified in the number of supporting exemplars. The present analysis offers a starting point for examining whether both properties may influence categorization through the same mechanism—category-specific sensitivity—pointing towards a unified account. People are generally sensitive to category prevalence and tend to favor majority categories (e.g., Maddox, 1995; Maddox & Bohil, 1998), a pattern compatible with exemplar models. At the same time, some prevalence-related phenomena challenge the similarity framework (see also Lewandowsky, 1995).

A particularly striking example is *prevalence-induced concept change* (Levari et al., 2018). When the prevalence of one category decreases over time, people paradoxically become more likely to assign ambiguous items to that now less prevalent category. Rather than contracting, the category appears to expand, even though fewer exemplars are encountered. This effect is difficult to reconcile with standard similarity-based predictions: prototype models disregard prevalence and its dynamics altogether, whereas exemplar models predict an opposite bias toward high-prevalence categories (cf. Albrecht & Spektor, 2025). Nevertheless, prevalence-induced concept change has proven robust across multiple manipulations, persisting even when participants were explicitly instructed or incentivized to maintain stable category boundaries (Levari et al., 2018). Importantly, Levari (2022) showed that adapting the range of the minority category can eliminate prevalence-induced concept change, pointing to a close connection between prevalence and variability.

Category-specific sensitivity provides a candidate mechanism for this effect. When a high-prevalence category $\mathcal{H}$ is encountered frequently, the cognitive system can form a relatively precise representation of its distribution, supporting high sensitivity to deviations from its prototype or exemplars. In contrast, a low-prevalence category $\mathcal{L}$ provides sparse evidence, potentially leading to uncertainty about its boundaries and lower sensitivity to deviations from the prototype or exemplars. Formally, if $c_{\mathcal{H}} \gg c_{\mathcal{L}}$, similarity-based models predict that ambiguous items are assigned to $\mathcal{L}$ despite its lower prevalence. Crucially, this account suggests a mechanism to unify the category variability effect and prevalence-induced concept change under a single principle: sensitivity scales with the degree to which a category is specified. High variability and low prevalence both reduce specification—one by stretching the category across feature space, the other by leaving it undersampled—and both may therefore yield low sensitivity. Future research should test this hypothesis directly (Levari, 2022), for example by examining whether experimental manipulations that decouple variability from prevalence produce additive or interactive effects on sensitivity.

## Conclusion

The present work demonstrates that the category variability effect—the tendency to assign ambiguous items into more variable categories—can be accommodated within the similarity framework, contrary to earlier claims (e.g., E. E. Smith & Sloman, 1994). Our simulations show that, of two proposed mechanisms, category-specific sensitivity provides a more robust and higher-performing account than response bias, particularly for prototype models. A compelling interpretation is that the mind calibrates its sensitivity to the precision with which each category is specified. This principle extends naturally to prevalence effects, pointing towards a unified similarity-based account in which sensitivity tracks the overall degree of category specification—whether determined by variability, prevalence, or both.

## References

Albrecht, R., Hoffmann, J. A., Pleskac, T. J., Rieskamp, J., & von Helversen, B. (2020). Competitive retrieval strategy

causes multimodal response distributions in multiple-cue judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(6), 1064–1090. doi: 10.1037/xlm0000772

Albrecht, R., & Spektor, M. S. (2025). Prevalence-induced concept change: Universal or context-dependent? implications for social psychology and AI cognition. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47).

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429. doi: 10.1037/0033-295X.98.3.409

Battleday, R. M., Peterson, J. C., & Griffiths, T. L. (2020). Capturing human categorization of natural images by combining deep networks and cognitive models. *Nature Communications*, *11*(1), 1–14. doi: 10.1038/s41467-020-18946-z

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, *29*(8), 1165–1175. doi: 10.3758/BF03206386

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*(2), 125–157. doi: 10.1016/0010-0277(94)90065-5

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin & Review*, *15*(2), 256–271. doi: 10.3758/PBR.15.2.256

Krefeld-Schwalb, A., Pachur, T., & Scheibehenne, B. (2022). Structural parameter interdependencies in computational models of cognition. *Psychological Review*, *129*(2), 313–339. doi: 10.1037/rev0000285

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. doi: 10.1037/0033-295X.99.1.22

Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1003–1021. doi: 10.1037/0278-7393.20.5.1003

Levari, D. E. (2022). Range-frequency effects can explain and eliminate prevalence-induced concept change. *Cognition*, *226*, 105196. doi: 10.1016/j.cognition.2022.105196

Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, *360*(6396), 1465–1467. doi: 10.1126/science.aap8731

Lewandowsky, S. (1995). Base-rate neglect in ALCOVE: A critical reevaluation. *Psychological Review*, *102*(1), 185–191.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* Oxford, England: John Wiley.

Maddox, W. T. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 288–301. doi: 10.1037/0278-7393.21.2.288

Maddox, W. T., & Bohil, C. J. (1998). Overestimation of base-rate differences in complex perceptual categories. *Perception & psychophysics*, *60*(4), 575–592. doi: 10.3758/bf03206047

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. doi: 10.1037/0033-295X.85.3.207

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. doi: 10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. doi: 10.1037/0278-7393.14.4.700

Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (Vol. 1, pp. 149–167). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*(3), 375–402. doi: 10.1007/BF03543066

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. doi: 10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review*, *5*(3), 345–369. doi: 10.3758/BF03208813

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79. doi: 10.1037/0033-295X.101.1.53

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), 924–940. doi: 10.1037//0278-7393.28.5.924

Perlman, A., Hahn, U., Edwards, D. J., & Pothos, E. M. (2012). Further attempts to clarify the importance of category variability for categorisation. *Journal of Cognitive Psychology*, *24*(2), 203–220. doi: 10.1080/20445911.2011.613818

Poggio, T., & Girosi, F. (1990). Networks for approximation and learning. *Proceedings of the IEEE*, *78*(9), 1481–1497.

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

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Lawrence Erlbaum Associates Hillsdale, NJ.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. doi: 10.1016/0010-0285(75)90024-9

Sakamoto, Y., Jones, M., & Love, B. C. (2008). Putting the psychology back into psychological models: Mechanistic versus rational approaches. *Memory & Cognition*, *36*(6), 1057–1065. doi: 10.3758/MC.36.6.1057

Seitz, F. I., Albrecht, R., von Helversen, B., Rieskamp, J., & Rosner, A. (2025). Identifying similarity-and rule-based processes in quantitative judgments: A multi-method approach combining cognitive modeling and eye tracking. *Psychonomic Bulletin & Review*, *32*, 1761–1775. doi: 10.3758/s13423-024-02624-y

Seitz, F. I., Jarecki, J. B., & Rieskamp, J. (2023). Modeling the category variability effect in an exemplar-similarity framework. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45). Retrieved from https://escholarship.org/uc/item/4z76q867

Seitz, F. I., Jarecki, J. B., & Rieskamp, J. (2026). Perceptual similarity mostly ignores within-category feature distributions: Evidence from computational modeling of human categorizations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Advance online publication. doi: 10.1037/xlm0001567

Seitz, F. I., Jarecki, J. B., Rieskamp, J., & von Helversen, B. (2025). Disentangling perceptual and process-related sources of behavioral variability in categorization. *Perspectives on Psychological Science*, *20*(3), 590–617. doi: 10.1177/17456916241264259

Seitz, F. I., von Helversen, B., Albrecht, R., Rieskamp, J., & Jarecki, J. B. (2023). Testing three coping strategies for time pressure in categorizations and similarity judgments. *Cognition*, *233*, 105358. doi: 10.1016/j.cognition.2022.105358

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. doi: 10.1126/science.3629243

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*(3), 278–304. doi: 10.1037//0096-3445.121.3.278

Smith, D. J., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1411–1436. doi: 10.1037/0278-7393.24.6.1411

Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory & Cognition*, *22*(4), 377–386. doi: 10.3758/BF03200864

Stewart, N., & Brown, G. D. (2004). Sequence effects in the categorization of tones varying in frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 416–430.

Stewart, N., Brown, G. D., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 3–11.

Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 893–907. doi: 10.1037//0278-7393.28.5.893

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640. doi: 10.1017/S0140525X01000061

Yang, D.-Y., & Worthy, D. A. (2026). Frequency effects in human category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *52*(2), 169–193.

Yang, L.-X., & Huang, T.-L. (2021). Exemplar account for category variability effect: Single category based categorization. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, pp. 2699–2705).

Yang, L.-X., & Wu, Y.-H. (2014). Category variability effect in category learning with auditory stimuli. *Frontiers in Psychology*, *5*, 1122. doi: 10.3389/fpsyg.2014.01122