ASSOCIATION FOR
PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2025, Vol. 20(3) 590–617

DOI: 10.1177/17456916241264259
www.psychologicalscience.org/PPS
Sage

Disentangling Perceptual and Process-Related Sources of Behavioral Variability in Categorization

Florian I. Seitz[1], Jana B. Jarecki[1], Jörg Rieskamp[1], and Bettina von Helversen[2]
[1]Center for Economic Psychology, Department of Psychology, University of Basel, and
[2]Department of Psychology, Faculty of Human and Health Sciences, University of Bremen

Abstract
People often categorize the same object variably over time. Such intraindividual behavioral variability is difficult to identify because it can be confused with a bias and can originate in different categorization steps. The current work discusses possible sources of behavioral variability in categorization, focusing on perceptual and cognitive processes, and reports a simulation with a similarity-based categorization model to disentangle these sources. The simulation showed that noise during perceptual or cognitive processes led to considerable misestimations of a response determinism parameter. Category responses could not identify the source of the behavioral variability because different forms of noise led to similar response patterns. However, continuous model predictions could identify the noise: Noisy feature perception led to variable predictions for central stimuli on the category boundary, noisy feature attention increased the prediction variability for stimuli differing from each category on another feature, and noisy similarity computation increased the variability for stimuli with moderate predictions. Measuring category beliefs in a continuous way (e.g., through category probability judgments) may therefore help to disentangle perceptual and process-related sources of behavioral variability. Ultimately, this can inform interventions aimed at improving human categorizations (e.g., diagnosis training) by indicating which steps of the categorization mechanism to target.

Keywords
variability, categorization, perception, cognition, computational modeling

The human mind often reacts differently to the same object over time. People may assign objects sometimes to one category and sometimes to another. For instance, people may categorize tomatoes as fruits or vegetables, turquoise as green or blue, and a consonant with intermediate voice-onset time as voiced or voiceless. Individuals' behavior can vary over time not only in categorizations (Ashby & Maddox, 1998; Beck et al., 2012; Stewart et al., 2002; Wyart & Koechlin, 2016) but also in other inferential choices (Gaissmaier & Schooler, 2008), preferential choices (Rieskamp et al., 2006), and quantitative judgments (Albrecht et al., 2020). Such behavioral variability can have various cognitive underpinnings, and an increase in behavioral variability (e.g., because of an experimental manipulation) could easily be interpreted as a change in cognitive strategy—yet people may have simply become less precise (Olschewski et al., 2018; Seitz, von Helversen, et al., 2023). To avoid such misinterpretations, one needs to break down behavioral variability and pinpoint its potential origins in the cognitive system. This article provides an overview of sources of behavioral variability in categorization and presents a simulation-based way to disentangle these sources in a cognitive categorization model.

Categorizing objects is fundamental for cognition. Much psychological research has sought to understand human categorizations, theorizing that people categorize objects by applying rules or similarity-based

Corresponding Author:
Florian I. Seitz, Center for Economic Psychology, Department of Psychology, University of Basel
Email: florian.seitz@unibas.ch

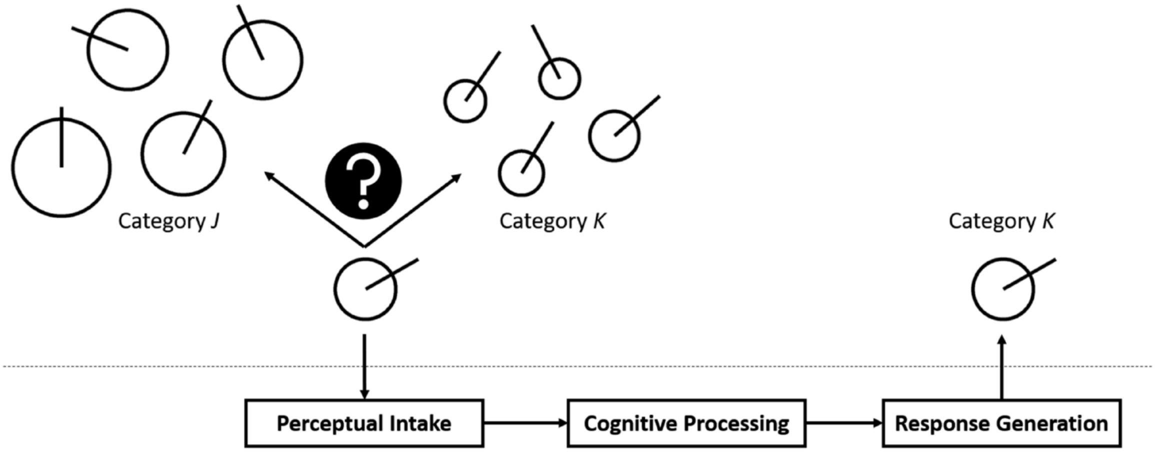

Fig. 1. Visualization of the categorization mechanism. In this example, objects are geometric figures with two features: a circle varying in size and a line varying in orientation. People first perceive the to-be-categorized object, then compute the evidence that it belongs to different categories, and finally select a category response. Each of these steps may be a source of behavioral variability.

processes (introduced in detail below; for an overview, see Ashby & Maddox, 2005; Palmeri et al., 2004; Richler & Palmeri, 2014; Serre, 2016). Over the years, many cognitive models have formalized these theories, and, broadly speaking, they successfully describe human category learning across a wide range of tasks. These tasks involve objects with features that are separable or integral (e.g., Nosofsky, 1986, 1987) and discrete or continuous (e.g., Cohen et al., 2001), categories that are artificial or naturalistic (e.g., Nosofsky et al., 2018) and include numerous or few exemplars (e.g., Maddox & Ashby, 1993), and environments in which categories do or do not overlap (e.g., Ell & Ashby, 2006) and are linearly or nonlinearly separable (e.g., Shepard et al., 1961). Although these models account for category learning, they often do not address the variability that people may show when transferring their category knowledge to new objects—across trials, a model makes the same prediction for the same object.

Yet behavioral variability is an important component of human categorization. During categorization, people perceive an object, process it to compute the evidence for different categories, and finally select a category response (Fig. 1; see also Wills & Pothos, 2012). Each of these steps can be a source of behavioral variability and thus of deviations from constant model predictions: During *perceptual intake*, people may perceive an object variably and differently from what serves as model input (e.g., Ashby & Lee, 1993). During *cognitive processing*, people may (compared with a model) compute category evidence in an imprecise and variable way (e.g., Wyart & Koechlin, 2016). Finally, during *response selection*, people may select a category other than the one that appears correct, which can be a pure error or the result of contextual factors limiting response precision (e.g., Seitz, von Helversen, et al., 2023). For each step of the categorization mechanism, the current work examines factors that may cause behavioral variability. We also present a simulation that disentangled these sources of behavioral variability in a similarity-based categorization model. Ultimately, our approach may not only inform categorization models but also help determine which step of the categorization mechanism one needs to target in interventions aimed at reducing behavioral variability in human categorizations.

The Importance and Difficulty of Identifying Behavioral Variability

Behavioral variability in categorization is ubiquitous but often undesirable. A physician who diagnoses the presence of skin cancer on the basis of medical imaging should not exhibit behavioral variability but consistently make the correct diagnosis. For interventions such as diagnosis training to be successful, it is important to know where the behavioral variability originates. However, this can prove difficult, notably because variability can be confused with biases in categorization data. Here, we focus on binary categorizations (analogous principles hold for multiclass classification) and define variability as the variance of a person's category responses for a stimulus and bias as the difference between the person's mean category response for this stimulus and the stimulus' true category label. Crucially, the mean and variance are interdependent: As the response variance increases, the

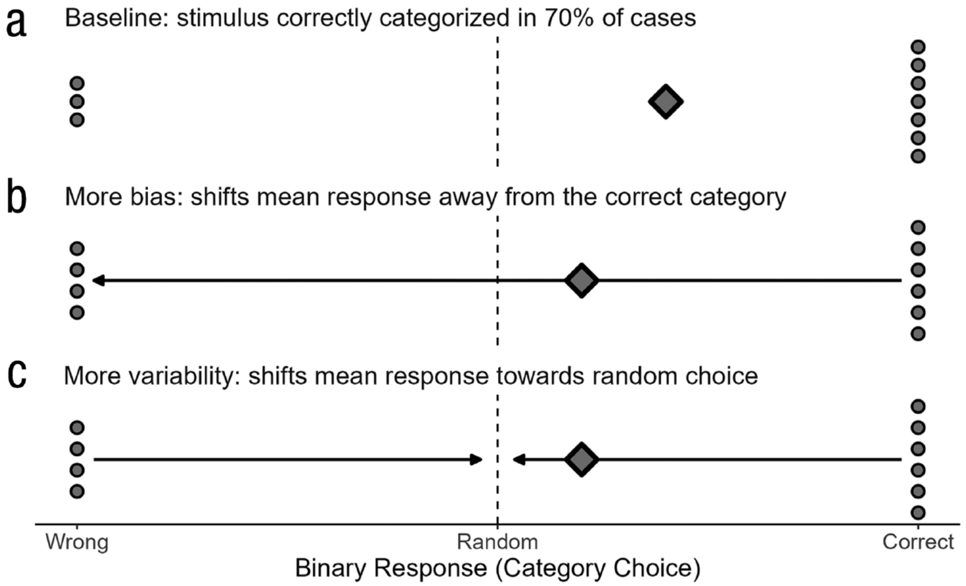
a
Baseline: stimulus correctly categorized in 70% of cases
b
More bias: shifts mean response away from the correct category
c
More variability: shifts mean response towards random choice
Wrong
Random
Correct
Binary Response (Category Choice)

Fig. 2. Interdependence of bias and variability in (binary) categorization data. The points show a person's repeated category responses for one object; the diamond represents the mean response. More bias increases the difference between the mean response and the correct category; more behavioral variability shifts the mean response toward random choice.

mean response shifts toward a random response (.5 for either category in the binary case). Figure 2 illustrates the relation between bias and variability: Relative to the upper baseline case, the mean response, represented as a diamond, is shifted further away from the correct category label in the case of more bias and closer to random choice in the case of more variability. As the arrows in Figure 2 show, these response shifts can go into the same direction, highlighting the interdependence of bias and variability.

From a psychological perspective, however, identifying behavioral variability seems possible only if one can distinguish it from a bias. Assigning an object sometimes to the wrong category can have various psychological reasons. For instance, noise in the form of random fluctuations during perceptual and cognitive processes might sometimes make the wrong category appear as the correct one. Alternatively, one might use a suboptimal (biased) categorization strategy that provides only limited evidence for the correct category. In both cases, a probabilistic response pattern emerges—the psychological conditions giving rise to it, however, are completely different.

The remainder of the article is structured as follows: We first provide an overview of major categorization theories and their sources of behavioral variability during perceptual intake and cognitive processing. We then present a simulation that formalized various such sources in a cognitive categorization model. One key result of the simulation was that continuous data helped to identify the source of behavioral variability. In the final section, we discuss how continuous data can be integrated into a categorization experiment and how this can also help to distinguish behavioral variability from behavioral biases.

Perceptual Categorization Theories and Behavioral Variability

This section introduces the main theories from the literature on perceptual categorization (Palmeri et al., 2004), which categorize objects on the basis of similarity (Medin & Schaffer, 1978; Nosofsky, 1986) or rules (e.g., Ashby & Gott, 1988; Nosofsky et al., 1994).

Similarity-based categorization theories assume that people assign an object to the most similar category (Nosofsky, 1986; Smith & Minda, 1998). This means people compute the object's similarity to representatives of each category, which can be individual exemplars or abstracted prototypes (Nosofsky & Zaki, 2002). The more similar an object is to the representatives of one category (relative to alternative categories), the more likely it belongs to that category. Thus, similar objects are grouped into the same category—an assumption that has received substantial support from categorization experiments (Minda & Smith, 2002;

Nosofsky, 1984, 1986, 1987, 1989; Nosofsky & Palmeri, 1997; Nosofsky & Zaki, 2002; Seitz, Jarecki, Rieskamp, 2023; Seitz, von Helversen, et al., 2023; Smith & Minda, 1998, 2000, 2002). More recent research has successfully applied similarity-based models to natural-object domains (Battleday et al., 2020; Meagher & Nosofsky, 2023; Nosofsky et al., 2022; Sanders & Nosofsky, 2020), providing further evidence that people categorize objects on the basis of their similarity to category representatives.

Rules, in turn, describe the conditions for category membership. Some rules specify the features needed for belonging to a category (Nosofsky et al., 1994); other rules specify the boundaries between categories and associate each resulting region of the feature space with a category response (Ashby & Gott, 1988). The two kinds of rules can be seen as complementary—one focuses on the content of the categories, and the other focuses on their borders (Kruschke, 2008). Rules can vary in complexity from being based on a single feature to combining several features in an intricate way. Some of these more complex rules can still be easily verbalized (e.g,. a conjunctive rule that combines two unidimensional rules with a logical AND), whereas others are almost impossible to verbalize (e.g., a diagonal decision rule in feature space). Categorization experiments have demonstrated that people can learn a variety of rules (Ashby & Gott, 1988; Ashby & Maddox, 1990, 1992; Ashby & Perrin, 1988; Maddox & Ashby, 1993; Maddox et al., 2004; Nosofsky & Palmeri, 1998; Nosofsky et al., 1994), and often these rules approximate optimal behavior that maximizes accuracy.

Although categorization models based on similarity or rule processes coincide in terms of perceptual intake (representing objects as vectors of feature values), they often differ in terms of response selection (deterministic or probabilistic). Deterministic models always select the most probable category according to the cognitive process used; probabilistic models, in turn, select responses in proportion to the probabilistic category beliefs (Maddox & Bohil, 2004). Response selection is usually formalized with a choice rule (Jarecki & Seitz, 2020; Nosofsky & Zaki, 2002; for an alternative, see Cavagnaro & Regenwetter, 2023). Whereas rules typically assume a deterministic choice rule (Ashby & Gott, 1988; Nosofsky et al., 1994) or a fixed error rate (trembling hand error; Scheibehenne et al., 2013), similarity-based categorization models use probabilistic choice rules that produce more or less variable behavior (Medin & Schaffer, 1978; Nosofsky, 1986; but see Ashby & Maddox, 1993; Maddox & Ashby, 1993). Similarity-based categorization models thus provide a particularly useful tool for analyzing behavioral variability; we now discuss potential sources of behavioral variability in similarity-based categorization models, focusing on perceptual intake and cognitive processing (for an overview, see Table 1).

Perceptual intake

Categorization variability can originate in perceptual processes, namely when people imprecisely perceive the feature values of a to-be-classified object (Alfonso-Reese, 2001; Ashby, 1992; Ashby & Lee, 1993; Maddox, 2001; Maddox & Bogdanov, 2000; Peterson et al., 2019). In Figure 1, for example, the object may be assigned to Category B in trials in which the object's circle size was correctly perceived or underestimated and to Category A in other trials in which the circle size was overestimated. In other words, a noisy perception of the feature "circle size" leads to a variable categorization of the to-be-classified object. This applies in particular to objects close to the category boundary, where misperception on a (diagnostic) feature can actually change categorizations; responses to objects far away from the boundary should remain largely unaffected by perceptual noise. Note that the perceptual imprecision above stems from noise across trials—on average, the features are correctly perceived. Alternatively, people can have a biased perception and systematically misperceive features in a certain direction. Such a perceptual bias leads to a shift in the category predictions, which at the response level, however, can also manifest itself as a change in behavioral variability because bias and variance are related in categorization data. In sum, a perceptual representation that differs from the physical features of the objects (in the form of a bias or noise) can lead to behavioral variability in categorizations.

Perceptual imprecision is supported by empirical evidence (Alfonso-Reese, 2001; Nosofsky, 1986; Petzschner et al., 2015) and can be caused by the object (e.g., physical noise such as the probabilistic emission of photons by a light source of constant intensity) or the organism (e.g., sensory noise such as a variation in pupil size; Ashby & Lee, 1993). When asked to adjust the features of a stimulus until they match those of a fixed reference stimulus, participants typically adjust the features insufficiently (Alfonso-Reese, 2001). The adjusted stimuli vary around the true reference stimulus, and interestingly for some features more than for others, suggesting that perceptual imprecision may be feature-specific. Perceptual imprecision is known for a wide range of features, including the size of lines (Stevens, 1960), rectangles (Krantz & Tversky, 1975), and circular stimuli (Nosofsky, 1986), angles (Petzschner & Glasauer, 2011), and color

Table 1. Overview of Potential Sources of Behavioral Variability in Categorization

Categorization step	Possible behavioral variability source	Example reference
Perceptual intake	Variable perception of feature values (e.g., additive noise)	Ashby and Lee (1993)
Cognitive processing		
Category representation	Variable use of exemplars	Nosofsky and Palmeri (1997)
Attention allocation	Variable attention to features	Kruschke (1992)
Similarity computation	Variable computation of similarity	Rodrigues and Murre (2007)
Multiple processes	Variable integration of processes	Erickson and Kruschke (1998)
Response selection	Errors (e.g., trembling hand) and contextual factors (e.g., time pressure)	Seitz, von Helversen, et al. (2023)

brightness (Nosofsky & Palmeri, 1996). In addition, people's perception depends on a feature's range of values (Petzschner & Glasauer, 2011) and is typically biased toward the mean of the presented feature values, with larger and more variable deviations for larger feature values (Petzschner et al., 2015).

Perceptual imprecision can affect cognitive inferences such as identifications and categorizations (Ashby, 2000; Maddox, 2001). Alfonso-Reese et al. (2002) found that categorization tasks, for which the maximally achievable accuracy decreases sharply with perceptual noise, are more difficult to learn than tasks that are unaffected by perceptual noise. Particularly difficult to learn are categories that are close to each other in space and have strongly correlated features (Ashby et al., 2020; Edmunds et al., 2015; Nosofsky et al., 2005), presumably because they are heavily affected by perceptual noise (Alfonso-Reese et al., 2002). Other research suggests that increasing expertise in a categorization task can make objects' perceptual representations more distinct (Goldstone, 1998; Goldstone et al., 2001; Palmeri et al., 2004). In particular for locations close to category boundaries, perceptual expertise allows for a high categorization accuracy by minimizing the probability that an object's percept lies on the wrong side of a decision boundary.

Although cognitive models of categorization often take an object's features directly as input, there are methods for incorporating perceptual imprecision. Multidimensional scaling can reveal perceptual biases by modeling people's perceptual feature space (e.g., Shepard, 1962a, 1962b). In addition to completing a categorization task, participants provide similarity relations for the same objects (e.g., pairwise similarity judgments or confusion errors during object identification; Nosofsky, 1986, 1989; Sanders & Nosofsky, 2020). Multidimensional scaling locates the objects in a feature space according to the provided similarity relations: Similar objects end up close to each other, and dissimilar objects end up further apart. Importantly, the inferred locations in space do not reflect the physical objects but instead represent their perceptual representations, which can then be used to model the categorization data. Note that for real-world objects psychologically interpreting the features that result from multidimensional scaling can be complicated because they need not correspond to the features of the physical space (Hebart et al., 2020; Izydorczyk & Bröder, 2023). Other research has addressed perceptual noise by modeling the perceived feature values of an object by a multivariate normal distribution, which contains a variance term for each feature (Ennis et al., 1988; Ennis & Johnson, 1993). This idea was central to the development of general recognition theory, which models the decision boundaries between categories for objects with a noisy percept (Ashby & Perrin, 1988; see also Ashby, 1992).

Cognitive processing

Behavioral variability can also originate during cognitive processing (Newell & Bröder, 2008). On the one hand, the human mind may perform a single process in a variable way (e.g., when computing the similarity to the category representatives). On the other hand, the human mind often combines multiple processes, and variability may arise if different weight is given to the individual processes across trials. We now elaborate on these factors, focusing on processes relying on the similarity to category representatives.

Category representations. There is a long-standing debate about whether the human mind represents categories by their individual members (*exemplars*; Nosofsky, 2011) or summary abstractions (*prototypes*; Minda & Smith, 2011). The exemplar theory claims that people store objects as exemplars in memory and assign a new object to the category to whose exemplars it is most similar. The prototype theory, in turn, summarizes each category by a prototype that corresponds to the category's central tendency (the mean value for each feature) and assigns a new object to the category with the most similar prototype. In both theories, the probability of choosing a category increases with an object's similarity to this category relative to its similarity to other categories (Nosofsky & Zaki, 2002), and this relationship is usually formalized by a variant of Luce's (1959) choice rule. In its simplest form, the choice rule computes the probability that object i belongs to category J as the similarity s_{iJ} of object i to category J divided by the summed similarity of i to all categories K, formally $s_{iJ} / \sum_K s_{iK}$. In an exemplar model, s_{iJ} equals the summed similarity of i to the category J members; in a prototype model, s_{iJ} equals i's similarity to the category J prototype. Although on average the exemplar theory is favored (e.g., Nosofsky, 1992), there is evidence that both exemplars and prototypes are important for category representations (Ashby & Maddox, 1993; Minda & Smith, 2002; Nosofsky, 1986; Nosofsky & Zaki, 2002; Smith & Minda, 1998; Zaki et al., 2003), and even today the debate is still relevant (e.g., Battleday et al., 2020; Nosofsky et al., 2022).

Recent research suggests that the human mind uses the two kinds of representations very flexibly. For instance, it has been suggested that prototypes and exemplars form the ends of a continuum that also contains intermediate stages in which multiple prototypes per category are formed (Verbeemen et al., 2007; Vanpaemel & Storms, 2008). The idea of an exemplar-prototype continuum also fits nicely with the finding that people sometimes shift from one representation to the other with increasing experience (e.g., Homa et al., 1981). Furthermore, the exemplar retrieval from memory does not have to be deterministic and include all exemplars (an *integrative retrieval*) but can also rely on a few probabilistically selected exemplars (a *competitive retrieval*; cf. Albrecht et al., 2020). Models implementing such a competitive retrieval (e.g., Nosofsky & Palmeri, 1997) in general formalize the probability that an exemplar is retrieved to be proportional to its similarity to the to-be-classified object. Because the competitive retrieval is probabilistic, however, different exemplars may be retrieved for the same object in different trials, which can result in behavioral variability.

Although competitive and integrative retrievals clearly differ from each other conceptually, they often make similar predictions. As Albrecht et al. (2020) noted, a competitive retrieval that samples only one exemplar makes the same average categorization predictions as an integrative retrieval. Simply put, the response prediction that results from averaging across all exemplars equals the average response prediction resulting from retrieving only one exemplar. More formally, consider a model that assigns any object i to the category to which one probabilistically retrieved exemplar belongs. The retrieval probability for any exemplar j equals the similarity s_{ij} between i and j, normalized by the summed similarity between i and the exemplars k from all categories $\sum_K \sum_{k \in K} s_{ik}$ to ensure that the probabilities sum up to 1. In this model, the probability that object i is assigned to category J equals the summed probability that an exemplar from category J is retrieved or, in other words, the normalized summed similarity to the exemplars from category J, formally $\sum_{j \in J} s_{ij} / \sum_K \sum_{k \in K} s_{ik}$. Yet this exactly corresponds to the response probability computed by Luce's choice rule for an integrative exemplar retrieval. In other words, the same behavioral variability may stem from a stable but probabilistic categorization belief based on an integrative retrieval or from a variable but deterministic categorization belief based on a competitive retrieval. In this case, the category responses cannot pinpoint the source of behavioral variability (however, process-tracing methods such as eye tracking may help shed light on the nature of the exemplar retrieval; see Rosner et al., 2022; Scholz et al., 2015; Seitz et al., 2024).

Similarity computation. Behavioral variability can also originate when computing an object's similarity to the category representatives. There are several approaches to similarity computation (for an overview, see Goldstone & Son, 2012; Roads & Love, 2023)—particularly well studied is the geometric approach, which assumes that two objects' similarity is based on their distance in feature space (cf. Nosofsky, 1986). Specifically, an object's feature values determine its coordinates in space, and the larger the distance d_{ij} between two objects i and j, the lower their similarity s_{ij}. The relation between similarity and distance is generally modeled by the negative exponential function $s_{ij} = \exp(-c \cdot d_{ij})$, known as Shepard's (1987) universal law of generalization, where $c \geq 0$ is a free similarity parameter reflecting the sensitivity to distances. Larger values for c make similarity decline more steeply with distance and can accentuate the distance differences among object pairs. In turn, d_{ij} is mostly formalized by the weighted Minkowski distance, leading to

$$s_{ij} = \exp\left(-c \cdot \left[\sum_{n=1}^{N} w_n \cdot |\, i_n - j_n \,|^r\right]^{\frac{1}{r}}\right), \qquad (1)$$

where i_n denotes object i's value on feature n, w_n is a free parameter representing the share of attention allocated to feature n (with $0 \leq w_n \leq 1$ and $\sum_n w_n = 1$), and r is the distance metric typically fixed to $r = 1$ (the city-block distance, used for objects with separable features) or $r = 2$ (the Euclidean distance for objects with integral features; see Nosofsky, 2011). Thus, Equation 1 computes a weighted sum of the feature value differences between objects i and j and transforms this distance into an inversely related similarity.

In the similarity framework of Equation 1, behavioral variability can for instance emerge during attention allocation, namely when the human mind distributes its attention to the object features differently across trials. Attention may fluctuate over time so that more attention is paid to one feature in some trials and to another feature in other trials (for a related idea on a probabilistic inclusion of features into the similarity computation process, see Lamberts, 1995, 1998; Lamberts & Brockdorff, 1997). If an object with two features differs equally from two categories but on a different feature, it is assigned to one category if one feature receives a lot of attention and to the other category if the other feature receives a lot of attention. Previous research has shown that attention can be exemplar-specific (Rodrigues & Murre, 2007; Sakamoto et al., 2004) and region-specific (Nosofsky & Hu, 2023) and that it may be reallocated as features are added or removed (Seitz, 2023). Furthermore, posterior distributions of the attention weights in Bayesian parameter estimations suggest that people do not distribute attention to the features in a perfectly constant way across trials (however, these results need to be handled with care because the data are aggregated across participants; M. Lee & Wetzels, 2010; Vanpaemel, 2009). Like in the distinction between noise and biases during feature perception, random attention fluctuations also need to be distinguished from directed shifts in attention: During category learning, attention can be selectively shifted to features that determine category membership, thereby maximizing categorization accuracy (called "selective attention"; Kruschke, 1992; Nosofsky, 1986, 1989). Adaptations in the distribution of attention can thus be expected as a sign of learning—but in the absence of learning, noisy attention may lead to behavioral variability.

Behavioral variability might also arise at a more general level during similarity computation. For instance, the human mind may possess various ways to compute similarity (Seitz, Jarecki, & Rieskamp, 2023; Tversky, 1977) and even combine different ways to compute similarity (Navarro & Lee, 2002). In a similar way, the distance sensitivity parameter c might vary over time, leading to behavioral variability. Psychologically, this means that in some trials large feature value differences are needed to make two objects dissimilar, whereas in other trials small differences suffice. In an exemplar model, a varying c can also be interpreted as a varying number of exemplars that determine categorization (similar to a competitive exemplar retrieval; see Collsiöö et al., 2023): Recall that larger values for c can accentuate differences among the distances. For instance, two object pairs with distances $d_{ij} = 1$ and $d_{ik} = 2$ have a similarity ratio of $s_{ij} / s_{ik} = 2.72$ under $c = 1$, but already $s_{ij} / s_{ik} = 7.39$ under $c = 2$. This means that with a larger c, more relative weight is given to the most similar exemplars, and thus fewer exemplars determine the category prediction.[1] Distance sensitivity has been suggested to be task-specific (Shin & Nosofsky, 1992), category-specific (Nosofsky & Johansen, 2000), or even exemplar-specific (Rodrigues & Murre, 2007; Schlegelmilch & von Helversen, 2020) and might thus fluctuate across trials, leading to a variable similarity between objects and, thereby, to variable categorizations.

Multiple cognitive processes. Behavioral variability can arise not only within a cognitive process but also when several processes are combined into a categorization strategy. Rules and similarity are often considered two distinct systems of category learning (cf. Ashby et al., 1998). Yet it is often not easy to distinguish between the two processes: A unidimensional rule makes similar predictions as a similarity-based process that strongly weights the feature that determines category membership according to the rule (Nosofsky et al., 1989). Accordingly, it is sometimes also claimed that rules and similarity represent the ends of a continuum (Newell et al., 2011; Pothos, 2005; Verguts & Fias, 2009). Irrespective of the theoretical viewpoint, if the human mind relies on both processes to perform categorizations, behavioral variability can emerge.

Decision-making theory has proposed two ways of how rules and similarity-based processes interact: *shifting* between the two processes (e.g., using a rule in one trial and similarity in another trial) and *blending* both processes within trials (e.g., averaging the predictions of both processes to form a hybrid response; see Bröder et al., 2017; Herzog & von Helversen, 2018). Empirical evidence shows that people shift between rules and similarity-based processes depending, among other things, on the task (Hoffmann et al., 2016; Mata et al., 2012; Trippas & Pachur, 2019; von Helversen et al., 2010, 2013). For example, people may rely on explicit rules to solve tasks determined by a single feature (rule-based tasks) but on implicit similarity processes in tasks that require the integration of information from multiple features (information-integration tasks; see Ashby & Ell, 2001; Ashby et al., 1998, 2020). Shifting between processes may also occur within a task across trials (see Rouder & Ratcliff, 2006; Thibaut et al., 2018). In a rule-plus-exception task, people generally use a (unidimensional) rule and store exceptions to this rule that can influence the classification of similar stimuli (Erickson & Kruschke, 2002; Nosofsky et al., 1994; Nosofsky & Palmeri, 1998). Finally, people can blend the two processes within trials (Albrecht et al., 2020; Bröder et al., 2017; Erickson & Kruschke, 1998). For instance, even in clear rule-based tasks, the classification of transfer stimuli still often also depends on their similarity to previously experienced exemplars (Allen & Brooks, 1991; Hahn et al., 2010; Lacroix et al., 2005; Thibaut & Gelaes, 2006).

Many more recent categorization models have combined rule and similarity-based processes (Anderson & Betz, 2001; Erickson & Kruschke, 1998; Love et al., 2004; Schlegelmilch et al., 2021). For instance, Erickson and Kruschke (1998) described a blending model that predicts categorizations on the basis of a weighted average between a similarity module and a rule module. Anderson and Betz (2001), in turn, assumed that in any categorization trial people will select between a similarity-based and a rule-based process. The process with the greatest utility will be selected with the highest probability, but occasionally people may choose the process with lower utility. Accordingly, people may apply different categorization processes across the repetitions of the same stimulus, which can lead to behavioral variability. Schlegelmilch et al. (2021) assumed a probabilistic learning of rules based on similarity processes. Behavioral variability is caused by differences in the strength of belief in the success of a rule. In other words, this model explains changes in behavioral variability across learning as the search for an appropriate strategy (high variability) that, once found, is executed consistently (low variability). Furthermore, the model learns partial rules that are applied only in specific contexts in which they are found to be successful.

Response selection and contextual factors

Generally, people aim to maximize categorization accuracy. To this end, they should consistently choose the category to which the object most likely belongs, formally denoted by the arguments of the maxima choice rule. Deviations from such an idealistic classifier are quickly imagined as unsystematic errors in response selection. However, this need not be the case: Particularly at the beginning of a categorization task, people match their response proportions to their probabilistic category beliefs and shift to deterministic responding only after gaining experience and finding a successful categorization strategy (Ashby & Maddox, 1992; Maddox & Bohil, 2004). Furthermore, behavioral variability can result from contextual factors that affect response precision or also the perceptual and cognitive processes previously discussed.

For instance, limiting people's cognitive capacities (e.g., through time pressure) makes their categorizations more variable without, however, affecting the modal responses (Lamberts, 1995; Seitz, von Helversen, et al., 2023). A cognitive model can capture this with changes in the value of a parameter that reflects overall response variability in a probabilistic choice rule (Seitz, von Helversen, et al., 2023). People thus seem to keep using the same categorization strategy even under time pressure, but execute it with less precision (in addition to perceptual and attentional changes; e.g., Lamberts & Brockdorff, 1997; Wills et al., 2015). These findings are in line with evidence from studies on preferential choices that have found that cognitive-capacity limitations primarily affect people's response variability and not the underlying preference itself (Olschewski et al., 2018; Olschewski & Rieskamp, 2021).

Similarly, the same object may be classified differently depending on the previously presented object, inducing sequence effects (Stewart et al., 2002; Yang & Wu, 2014). Interestingly, this seemingly random response behavior can be formalized by a simple cognitive strategy according to which the probability of choosing the same category as in the previous trial is based on the similarity between the stimuli of the two trials (Stewart et al., 2002). Such sequence effects might have their origin in perceptual intake because the perception of a stimulus is influenced by stimuli presented on previous trials (Jones et al., 2006). Relatedly, people sometimes try to detect patterns concerning category membership in the sequence of presented objects and to this end select categories proportionally to their evidence (Gaissmaier & Schooler, 2008).

Finally, behavioral variability may be larger in more difficult categorization tasks, such as tasks with a large category overlap (addressed in the subsequent simulations). Category overlap makes a categorization task probabilistic (the feedback a category learner receives is not consistent across the repetitions of a stimulus) and thereby promotes probabilistic category responses as well (Jarecki et al., 2018; Little & Lewandowsky, 2009b). Indeed, people have been found to respond by matching their category responses to the actual category probabilities (probability matching; e.g., Little & Lewandowsky, 2009a). Moreover, exemplar-specific rewards for correct answers can affect behavioral variability (Maddox & Bohil, 1998), for example, by altering the distance sensitivity (Schlegelmilch & von Helversen, 2020), further highlighting the importance of contextual factors.

Summary

This section gave an overview of potential sources of behavioral variability in categorization. During perceptual intake, the translation of physical into perceptual feature values can be noisy (Ashby & Lee, 1993). This may lead to behavioral variability, for example, because the object's percept crosses a category boundary in some trials. Similarly, cognitive processing may be subject to variability either within a single process (e.g., during attention allocation or similarity computation; Nosofsky, 1984; Seitz, Jarecki, & Rieskamp, 2023) or when combining multiple processes (e.g., combining similarity-based and rule-based processes; Erickson & Kruschke, 1998). Finally, behavioral variability can also arise from errors in response selection and contextual factors, such as time pressure, that limit the precision with which one can perform a categorization process.

Model Simulation

To disentangle the different sources of behavioral variability in categorizations, we ran simulations with an exemplar model (Nosofsky, 1986). The model categorizes objects on the basis of a variant of Luce's choice rule; we implemented the softmax choice rule. Given two categories J and K, the probability $\Pr(J \mid i)$ that object i is assigned to category J is

$$\Pr\left(J \middle| i\right) = \frac{\exp\left(\tau \cdot \sum_{j \in J} s_{ij}\right)}{\exp\left(\tau \cdot \sum_{j \in J} s_{ij}\right) + \exp\left(\tau \cdot \sum_{k \in K} s_{ik}\right)}, \qquad (2)$$

where $\tau \geq 0$ is a free response determinism parameter. A large value for τ means deterministically selecting the more similar category; a small value for τ close to 0 shifts response proportions toward random choice. For instance, an object i with a summed similarity of $\Sigma_{j \in J} s_{ij} = 1.5$ to category J and a summed similarity of $\Sigma_{k \in K} s_{ik} = 0.5$ to category K is assigned to category J with $\Pr(J \mid i) = .73$ when $\tau = 1$ but with $\Pr(J \mid i) = .88$ when $\tau = 2$. The similarity s_{ij} between objects i and j is as defined in Equation 1.

$$s_{ij} = \exp\left(-c \cdot \left[\sum_{n=1}^{N} w_n \cdot \left| i_n - j_n \right|^r \right]^{\frac{1}{r}} \right)$$

The distance norm r was fixed to $r = 2$ for the simulations.[2] Thus, the model contained the following free model parameters: $N - 1$ attention weights w_n (with $0 \leq w_n \leq 1$ and $\sum_n w_n = 1$), the distance sensitivity $c \geq 0$, and the response determinism $\tau \geq 0$.

We ran simulations for two frequently used category structures—rule-based and information-integration structures (e.g., Ashby et al., 1998; Ell & Ashby, 2006; Maddox et al., 2003, 2004). In rule-based structures, category membership is determined by a single feature, and the optimal decision boundary is perpendicular to that feature. In information-integration structures, in turn, category membership is determined by multiple features conjointly, and the optimal decision boundary lies at an oblique angle to the features in space. Our simulations were based on the binary categorization tasks of Ell and Ashby (2006), who used rule-based and information-integration structures for objects with two features with various degrees of overlap between the two categories. The category overlap describes the extent to which a categorization task is probabilistic, in the sense that a feature value combination may be associated with multiple categories, and is therefore a useful proxy for the task difficulty. Ell and Ashby implemented five category overlap conditions (see their Table 1)—our simulations adopted the three intermediate conditions, which we labeled "low," "medium," and "high" overlap. As in Ell and Ashby, each category is represented by a bivariate normal distribution, defined by the two features' means, variances, and covariance. Our simulations abstracted away from the visualization of the features and pertain to any kind of objects with two features.[3] We rescaled the feature values by a factor of .1 to ensure they lie in a typical range for exemplar models. The resulting values for the two categories (labeled "left" and "right" on the basis of their mean value on Feature 1) are shown in Table 2 and visualized in Figure 3.

For each of the six category structures, 100 exemplars per category were sampled according to the category distributions specified in Table 2. Parameter values were also randomly sampled for the exemplar model from uniform distributions (i.e., the attention weight w_1 between 0 and 1, the distance sensitivity c between 0 and 2, and the response determinism parameter τ between 0 and 2).[4] Given the category exemplars and the parameter combination, the exemplar model made predictions for 25 fixed transfer stimuli that were evenly distributed across the feature space and repeated 20 times each (the dots in Fig. 3). Some of these predictions were noise-free; others contained a form of perceptual intake or cognitive-processing noise that was randomly sampled in each trial. We simulated noise during perceptual intake (called "noisy perception") by assuming a noisy representation of the feature values of the to-be-classified object i. Specifically, for each feature n the perceived feature value $\tilde{\imath}_n$ was equal to the true value i_n plus some feature-specific noise ϵ_n, $\tilde{\imath}_n \sim i_n + \epsilon_n$. We also simulated three potential sources of

Table 2. Bivariate Normal Distribution Parameters Used to Create the Learning Stimuli for the Rule-Based and Information-Integration Tasks at Different Levels of Category Overlap

	Means						
	Left category		Right category		Variances		
Overlap	Feature 1	Feature 2	Feature 1	Feature 2	Feature 1	Feature 2	Covariance
Rule-based category structures							
Low	8.53	11.00	13.47	11.00	0.51	2.75	0
Medium	9.80	11.00	12.20	11.00	0.51	2.75	0
High	10.45	11.00	11.55	11.00	0.51	2.75	0
Information-integration category structures							
Low	9.25	12.75	12.75	9.25	1.63	1.63	1.12
Medium	10.15	11.85	11.85	10.15	1.63	1.63	1.12
High	10.61	11.39	11.39	10.61	1.63	1.63	1.12

Note: The parameter values are based on the medium-low, medium, and medium-high conditions from Ell and Ashby (2006); for simplicity, we call the conditions "low," "medium," and "high," respectively. Variances and covariances apply to both categories.

cognitive-processing noise: attention allocation (called "noisy attention"), distance sensitivity (called "noisy sensitivity"), and similarity computation (called "noisy similarity"). The attention $\widetilde{w}_1$ allocated to Feature 1 in any trial was equal to the true attention weight w_1 plus ϵ_w, $\widetilde{w}_1 \sim w_1 + \epsilon_w$ (with $\widetilde{w}_2 = 1 - \widetilde{w}_1$). Similar error terms could affect the distance sensitivity ($\widetilde{c} \sim c + \epsilon_c$) and the similarity ($\widetilde{s}_{ij} \sim s_{ij} + \epsilon$). In all cases, the noise (i.e., ϵ_n, ϵ_w, ϵ_c, and ϵ_s) was modeled as stemming from a uniform distribution whose upper and lower boundaries were set at ± 1 *SD* of the distribution of the true values. Formally the trial-specific noise is given as

$$\begin{aligned}\epsilon_n &\sim \mathrm{U}\left(-\sigma_n, +\sigma_n\right),\\ \epsilon_w &\sim \mathrm{U}\left(-\sigma_w, +\sigma_w\right),\\ \epsilon_c &\sim \mathrm{U}\left(-\sigma_c, +\sigma_c\right),\\ \epsilon_s &\sim \mathrm{U}\left(-\sigma_s, +\sigma_s\right),\end{aligned} \tag{3}$$

with the standard deviations σ_n of the normally distributed feature values i_n equaling the square root of the variances in Table 2 and the standard deviations of the uniformly distributed parameters being $\sigma_w = .29$ and $\sigma_c = .58$. For simplicity, the similarities ranging from 0 to 1 were also assumed to be uniformly distributed, resulting in $\sigma_s = .29$.

For each of the six category structures, 500 simulation iterations were run, and at each iteration, new exemplars and parameter values were sampled. Given the exemplars and parameter values, the simulation applied one of the five forms of noise (i.e., no noise, noisy perception, noisy attention, noisy sensitivity, or noisy similarity) when making predictions for the transfer stimuli per Equation 2. We analyzed how trial-specific noise affected (a) the parameter recovery when the exemplar model was fit to binary category responses R_i sampled from the category predictions $\Pr(Right \mid i)$ of Equation 2, $R_i \sim \mathrm{Bernoulli}\left(\Pr(Right \mid i)\right)$; (b) the simulated binary category responses R_i; and (c) the continuous category predictions $\Pr(Right \mid i)$. The main text aggregates the results across the category overlap conditions; the results for all individual conditions are reported in Appendix A. All code can be found on https://osf.io/fbved/.

Parameter recovery

The model parameters were well recovered from the binary category responses in the noise-free case. Figure 4 shows the true and estimated values for the attention weight w, the distance sensitivity c, and the response determinism τ. In the left outer column without noise, the points are very close to the diagonal, indicating perfect recovery—especially for w and c (with the correlations between the true parameter values and the estimated values being $r_w = .92$ and $r_c = .89$ for the rule-based category structures and $r_w = .96$ and $r_c = .90$ for the information-integration category structures; see Appendix B). Parameter τ was recovered well at low values but less so at higher values because at some point a higher τ made the model only marginally more deterministic ($r_\tau = .82$).

Adding noise considerably worsened the parameter recovery. Although the estimated values for the attention weight parameter w were still quite close to the true values (see Fig. 4), the distance sensitivity parameter c and the response determinism parameter τ were often misestimated. Accordingly, the correlations between the true and estimated parameter values were

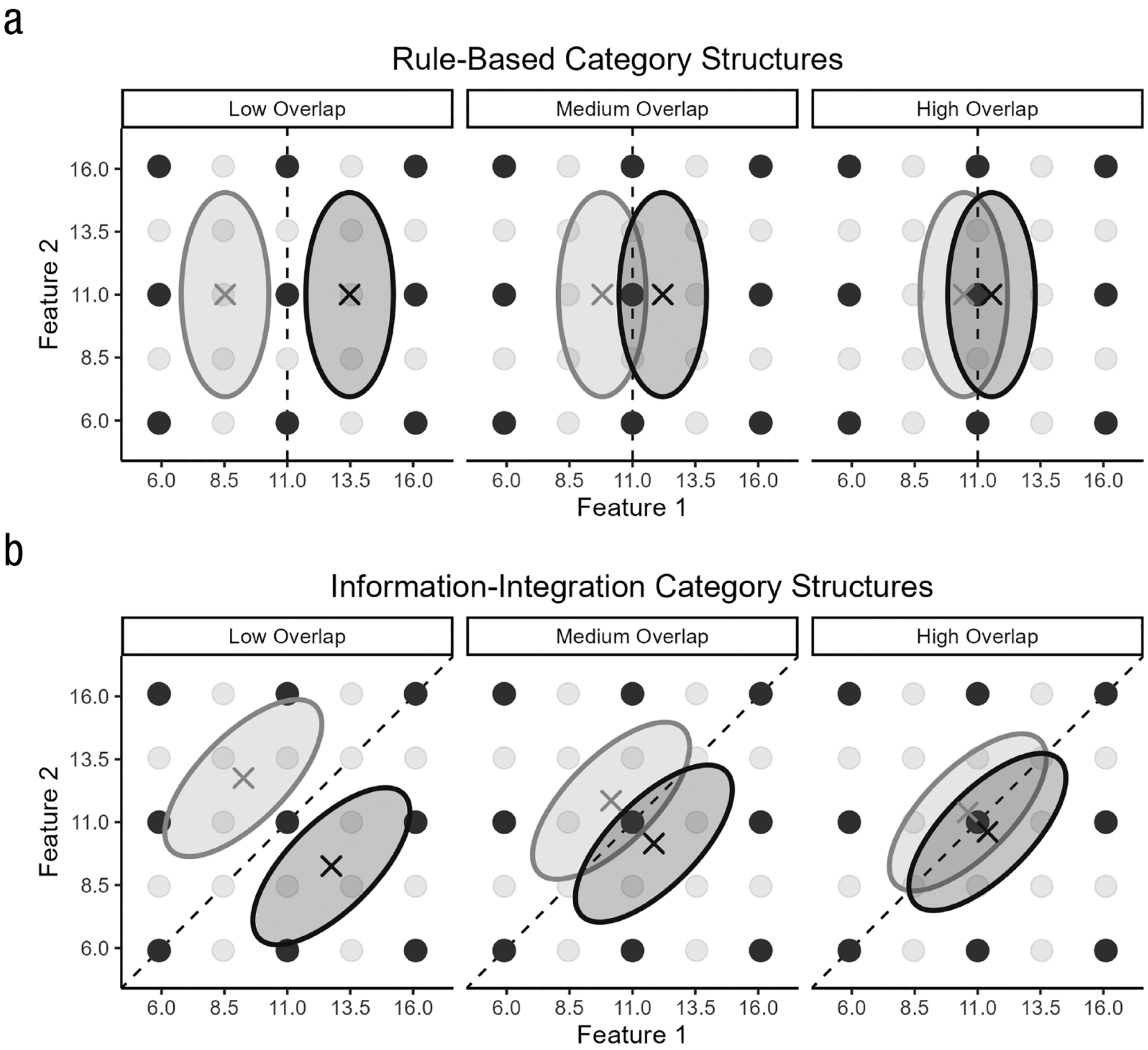

Fig. 3. Visualization of the (a) rule-based category structures and (b) information-integration category structures with a low, medium, or high category overlap. Each category is represented by the mean feature values (*x*) and the 95% density ellipse (left category in light gray; right category in dark gray). The 25 dots show the transfer stimuli that were fixed throughout the simulations (for clarity, Figs. 5 and 6 show the results only for the nine highlighted stimuli). Dashed lines show the category boundary maximizing accuracy.

larger for *w* (ranging from .85 to .95; see Appendix B) than for *c* (ranging from .76 to .91) and for τ (ranging from .37 to .65). In particular, τ was severely underestimated, especially when noise occurred during feature perception or similarity computation (see Fig. 4; for additional goodness-of-fit measures, see Appendix B).

These results show that the different parameters reflected the variability in the data to different degrees. The attention weight parameter *w* is feature-specific: It determines which feature receives how much weight but in general cannot shift all predictions to more or less random choice. Accordingly, the estimation of *w* was mostly unaffected by the variability present in the data. The distance sensitivity *c* and the response determinism τ, in turn, are global parameters that shift all predictions to random choice when converging to 0. Accordingly, the underestimation of τ suggests that the parameter tried to pick up the variability in the data stemming from earlier steps of the categorization mechanism. This may be unproblematic for pure response prediction in many cases. However, it impedes interpreting the parameters psychologically because the behavioral variability is attributed too much to response

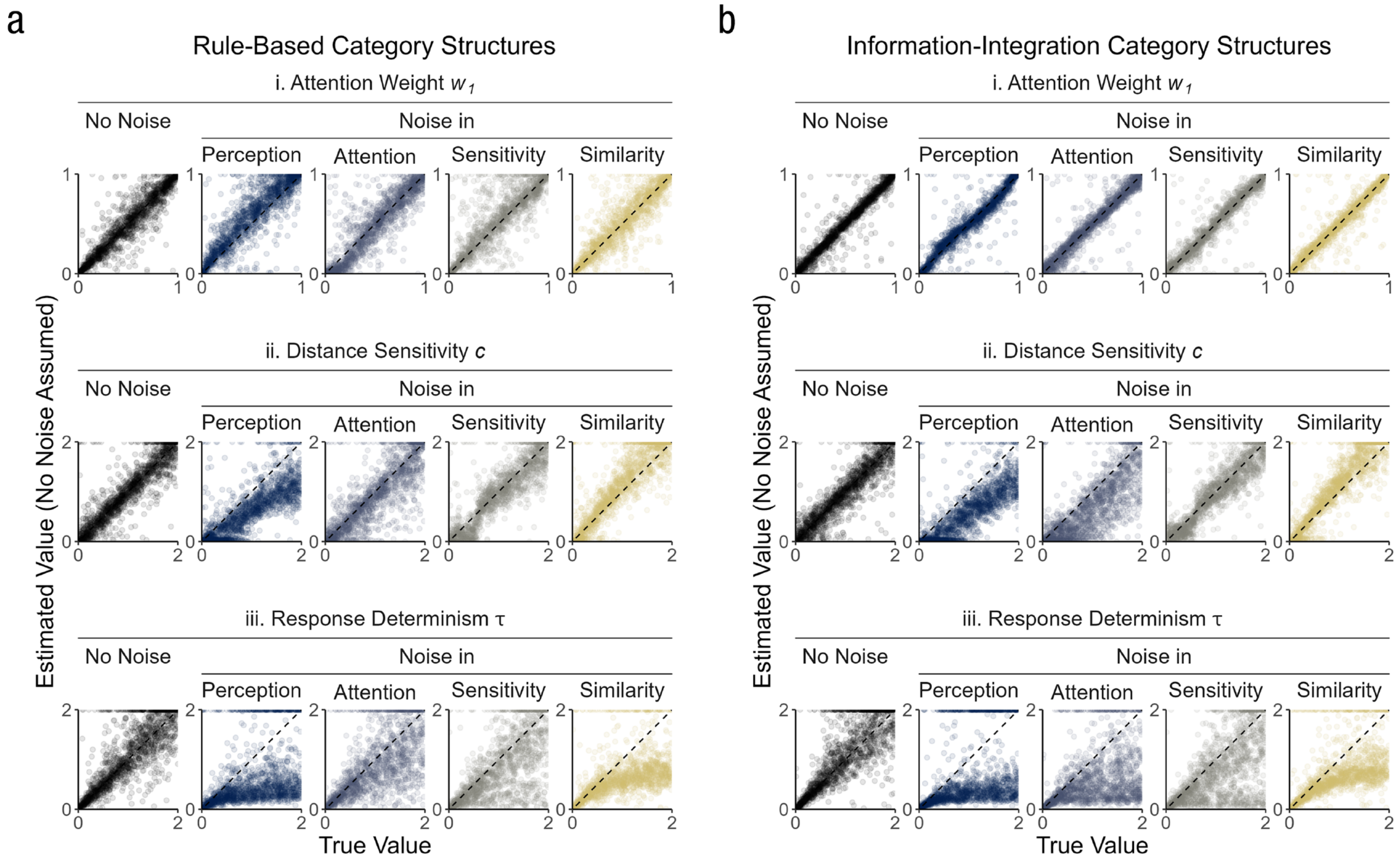

Fig. 4. Parameter recovery for the attention weight w_1, the distance sensitivity c, and the response determinism τ under different forms of noise in the (a) rule-based category structures and (b) information-integration category structures. The x-axis shows the true parameter values, the y-axis shows the estimated values, and the dashed diagonal represents a perfect recovery of the parameters from the simulated binary category responses.

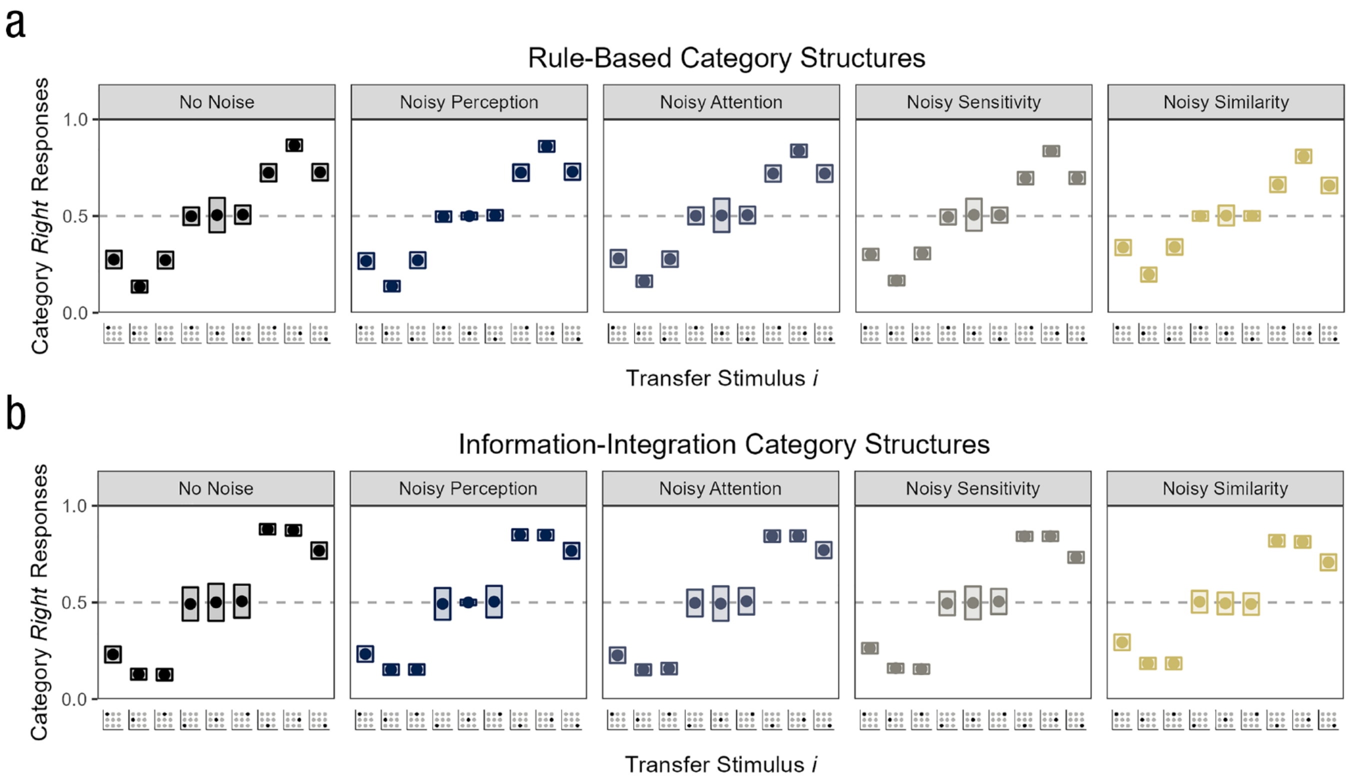

Fig. 5. Simulated binary category responses under different forms of noise for the (a) rule-based category structures and (b) information-integration category structures. For each transfer stimulus on the *x*-axis (shown are the nine stimuli highlighted in Fig. 3), the *y*-axis shows the mean and the variance of the response proportions for the right category across simulation iterations.

selection, disregarding perceptual intake and cognitive processing. In other words, one might not be able to identify the variability source correctly and on top of it misinterpret parts of the cognitive process (in particular the distance sensitivity *c*).

Binary category responses

If noise deteriorates the parameter recovery, it must also affect the binary category responses from which the parameters are estimated. Determining the source of the variability from the category responses, however, is difficult. Figure 5 shows that the response proportions for the individual transfer stimuli were on average very similar across the different forms of noise (for the results in the individual overlap conditions, see Fig. A1). In other words, regardless of what form of noise was administered, the model chose categories fairly deterministically for the transfer stimuli close to a category (the three transfer stimuli on the far left and far right of Fig. 5) and randomly for the transfer stimuli on the category boundary (the three stimuli in the middle of Fig. 5).

The response proportions also varied across the iterations of the simulation in a similar way for the different forms of noise. The rectangles in Figure 5 show that the response proportions varied notably only for the stimuli on the category boundary, and this variance was similar for the conditions with no noise, noisy attention, noisy sensitivity, and noisy similarity (e.g., for the central transfer stimulus, the mean variance across category structures was .09 in the case of no noise, .09 for noisy attention, .08 for noisy sensitivity, and .05 for noisy similarity). In the case of noisy perception, the variance for the central transfer stimulus was lower (M = .02), meaning that the category response proportion was about .50 for either category in every iteration of the simulation. One reason for this is that a noisy perception of the central transfer stimulus strongly influences which category contains the most similar exemplars and is thus selected, leading to response proportions close to .50 within each simulation iteration (recall that the noise was randomly sampled in each trial). For the remaining transfer stimuli on the category bound, the category containing the most similar exemplars is affected less by a noisy feature perception because there are less exemplars in these regions of the feature space. This leads to more deterministic response proportions within iterations (and, when aggregated across iterations, to mean response proportions close to .50, because each of the two categories contains the most similar exemplars in half of the iterations). Despite these subtle differences, Figure 5 clearly suggests that the category responses failed to reveal the source of the variability—both the response proportions within iterations and the variance of the response proportions across iterations shared a similar profile for different noise forms.

Continuous category predictions

One way to gain deeper insights into the source of the variability in a categorization experiment might be to include some form of continuous data that has a finer degree of resolution than the binary category responses discussed above. This builds a bridge to the related, extensive literature on quantitative judgment (Hoffmann et al., 2013, 2014; Juslin et al., 2003, 2008; Karlsson et al., 2007; Scheibehenne et al., 2015; von Helversen et al., 2014) that has already shown that a continuous criterion can be helpful in explaining various aspects of response distributions (e.g., Albrecht et al., 2020; Collsiöö et al., 2023; Sundh et al., 2021). For instance, Albrecht et al. (2020) rigorously compared different quantitative judgment models for their ability to predict response distributions, including similarity-based models with a competitive or integrative exemplar retrieval, a rule-based model, and two blending models. Interestingly, the competitive exemplar retrieval can predict multimodal response distributions within stimuli, because in each trial in which the stimulus is presented different exemplars may be retrieved. In contrast, the deterministic integrative retrieval can predict only unimodal response distributions. In their experiments, individual participants displayed multimodal response distributions in line with the predictions from a blending model combining a competitive exemplar retrieval with a rule-based process. The multimodality stemmed from the competitive exemplar retrieval; blending by itself leads to unimodal response distributions (compared with trial-wise shifting, which also can lead to multimodal response distributions).

Looking at continuous category predictions could have similar benefits in our simulation, and probabilistic categorization models are good candidates to test whether different forms of noise affect the continuous predictions differently. Figure 6 shows the exemplar model's continuous category predictions per Equation 2 for our simulation (for the results in the individual overlap conditions, see Fig. A2). Similar to the category responses, the mean predictions for the different transfer stimuli were very similar across the noise conditions (i.e., close to .5 for the stimuli on the category boundary and relatively deterministic for the remaining transfer stimuli). This makes sense given that the category responses were directly sampled from the continuous predictions. In contrast, the variability of the predictions within simulation iterations (the rectangles in Fig. 6) revealed the source of the trial-specific noise. Compared with the condition without noise (in which the within-iteration variability was of course zero), the four forms of noise increased the prediction variability in distinct ways.

Noisy perception resulted in variable predictions for the transfer stimuli on the category boundary, and in particular for the central transfer stimulus (as shown in Figure 6, the median standard deviation was .46 for both the rule-based and information-integration structures). This is because the predictions for transfer stimuli located close to the category boundary and in dense regions of the feature space are particularly easily affected by noise during feature perception. Noisy attention increased the prediction variability only rarely, namely for the transfer stimuli at the ends of the category boundary in information-integration category structures, $Mdn(SD) = .07$. This is because only these transfer stimuli differ from the exemplars of one category on one feature and from the exemplars of the other category on the other feature (see Fig. 3). For instance, for the lower left stimulus in an information-integration structure, more attention to Feature 1 increases its similarity to the left category, whereas more attention to Feature 2 increases its similarity to the right category. In contrast, the central transfer stimulus does not become more similar to either category with noisy attention because on each feature it is similar to some exemplars from both categories.

Unlike noisy perception and noisy attention, noisy sensitivity and noisy similarities made the predictions for most transfer stimuli more variable (see Fig. 6), likely because the two noise sources affect a later, more general step in the similarity computation. In the case of noisy sensitivity, the transfer stimuli with $\Pr(Right \mid i)$ around .25 or .75 were particularly affected because medium-deterministic predictions can mathematically change the most with fluctuating distance sensitivity, for example, the upper left stimulus with $Mdn(SD) = .10$ in the rule-based structures and $Mdn(SD) = .12$ in the information-integration structures. In contrast, noisy sensitivity affected predictions $\Pr(Right \mid i)$ close to 0, .5, or 1 less because its effect was outweighed by a clear distance pattern to the exemplars (i.e., equally large to the exemplars from both categories in the case of predictions

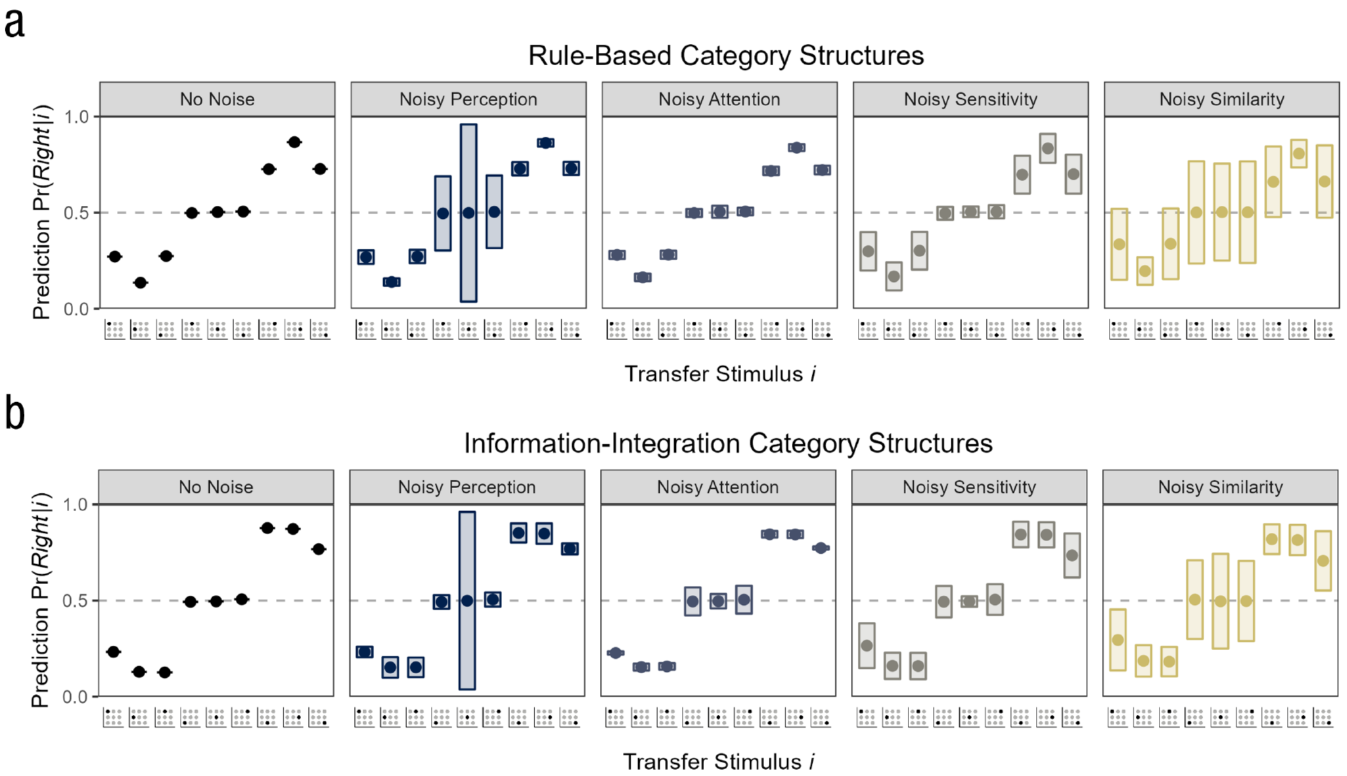

Fig. 6. Simulated continuous category predictions under different noise forms for the (a) rule-based category structures and (b) information-integration category structures. For each transfer stimulus on the *x*-axis, the *y*-axis shows the mean prediction Pr(*Right* | *i*) and the standard deviation of the predictions within simulation iterations (aggregated across iterations with the median).

around .5 and much larger to the exemplars from one category in the case of predictions close to 0 or 1). For noisy similarities, the transfer stimuli with Pr(*Right* | *i*) around .5 were affected the most, for example, the central stimulus with *Mdn*(*SD*) = .24 in the rule-based structures and *Mdn*(*SD*) = .25 in the information-integration structures. In general, the prediction variability is higher for noisy similarity than for noisy sensitivity; however, this of course also depends on the range in which the noise for these two sources may vary.

Summary

Our simulation suggests three main results. First, noise during perceptual and cognitive processes can substantially bias the parameter estimates of a model that does not take into account the noise during fitting on the binary category responses (for an analogous parameter recovery on the continuous category responses, see Appendix C). We observed an underestimation of the response determinism parameter τ, which attempted to capture the behavioral variability originating in earlier perceptual and cognitive processes, and of the distance sensitivity c, which characterizes the similarity computation process. Thus, in addition to being agnostic about the source of behavioral variability, the parameter estimates might lead to wrong inferences about the cognitive process when interpreted without care. Second, the noise cannot be identified by the binary category responses; both the response proportions that can be collected within an experiment and the variance of the response proportions across experiments show a similar pattern for different forms of noise. However, continuous data may provide a remedy to this problem, perhaps even within a single experiment, because different forms of noise lead to distinct variability patterns in the probabilistic category predictions. Specifically, noisy perception leads to variable predictions for central stimuli close to the category boundary, noisy attention increases the variability for stimuli differing from each category on another feature, and more general forms of noise during similarity computation (operationalized as a noisy distance sensitivity or noisy similarities) increase the variability for most stimuli with moderate predictions. In what follows, we discuss what this implies for categorization research and how one could integrate continuous data into a categorization experiment.

Integrating Continuous Data Into a Categorization Experiment

Our simulation suggests that integrating continuous data into categorization experiments might be helpful in identifying the source of behavioral variability and characterizing the cognitive process more accurately. However, where could one get continuous data in a categorization task that by definition requires discrete responses? One idea is a graded categorization task that measures the probability of belonging to a category or the inclusion of other similar continuous response measures (e.g., J. C. Lee et al., 2019; Lovibond et al., 2020). For example, after categorizing a stimulus, participants could assess how confident they are in their response (Balakrishnan & Ratcliff, 1996; Estes, 2004; Sieck & Yates, 2001). Such confidence judgments might reflect the continuous predictions that are output by a probabilistic categorization model: Higher confidence judgments may correspond to more deterministic predictions in favor of the chosen category (for a discussion, see Estes, 2004). Alternatively, one might not even need to expand the categorization task, but instead look at participants' response times (for examples on modeling people's response times in categorization, see Lamberts, 2000; Nosofsky & Little, 2010; Nosofsky & Palmeri, 1997). A shorter response time may reflect a higher confidence in one's response and thus correspond to a deterministic category belief.

To exemplify how category probability judgments, confidence judgments, and response times might help identify the source of behavioral variability, reconsider the central transfer stimulus from our model simulation. In the case of a noisy feature perception, the stimulus was sometimes highly likely to belong to one category and sometimes highly likely to belong to the other category, resulting in a substantial prediction variability across trials. This pattern may have been reflected in consistently high category probability and confidence judgments and fast response times (after all, the prediction in each trial was deterministic), but sometimes one category was chosen and sometimes the other. In contrast, noisy attention did not lead to any noteworthy predictive variability because no category was favored over the other under any attention distribution. This would correspond to the same category response proportions as before (i.e., close to 50% for each category) but to consistently low confidence judgments and slow response times (one is unsure about the categorization in every trial). Analogous demonstrations held for the other transfer stimuli, highlighting the potential benefits of including continuous data in categorization.

Zooming out, integrating continuous data can also help in disentangling behavioral variability from a directional bias (i.e., a systematic difference of a person's mean response from the true category label). Recall the introductory categorization example showing that increases in bias and variability lead to qualitatively similar response shifts (a shift away from perfect categorization and toward random choice). This was because in categorization data, the mean and the variability of the responses are interdependent. In contrast, in continuous responses, bias and variability are independent of each other and can readily be distinguished (see Fig. 7): More bias increases the difference between the mean response and the true category label; more variability increases the dispersion of the responses. Furthermore, a continuous measure also distinguishes between different response distributions for a given bias, such as unimodal distributions with a smaller or larger variability (see Fig. 7) or even multimodal distributions (see Albrecht et al., 2020). This may be particularly useful in probabilistic categorization tasks in which continuous responses such as category probability judgments could discriminate between a bias (the mean response deviates from the true category probability) and variability (the responses vary around the true category probability in a unimodal or multimodal way).

Whereas our approach of including continuous data to pinpoint the source of behavioral variability seems fruitful from a theoretical side, it also faces some practical challenges. First, many stimulus repetitions may be needed to get reliable estimates of the variability on the measure that aims to reflect continuous category predictions. Statistically, the variance of the estimates of behavioral variability decreases with the number of repetitions of a transfer stimulus and tends toward 0 as n approaches infinity. Furthermore, the larger the true behavioral variability, the more repetitions are needed get the same precision in estimation (see Casella & Berger, 2021, Example 7.3.). Thus, in particular when there is much behavioral variability, many repetitions are needed to properly identify it.

Second, designing an experiment such that it is able to elicit the source of behavioral variability is not an easy undertaking. As Figure 6 shows, the transfer stimuli differed in their ability to disentangle different forms of noise. For instance, one promising candidate for uncovering perceptual noise is the central transfer stimulus on the category boundary. However, this transfer stimulus might provide less insights for other endeavors such as model comparisons because most if not all sensible models will categorize this stimulus randomly. A solution might be to include this stimulus as a filler stimulus

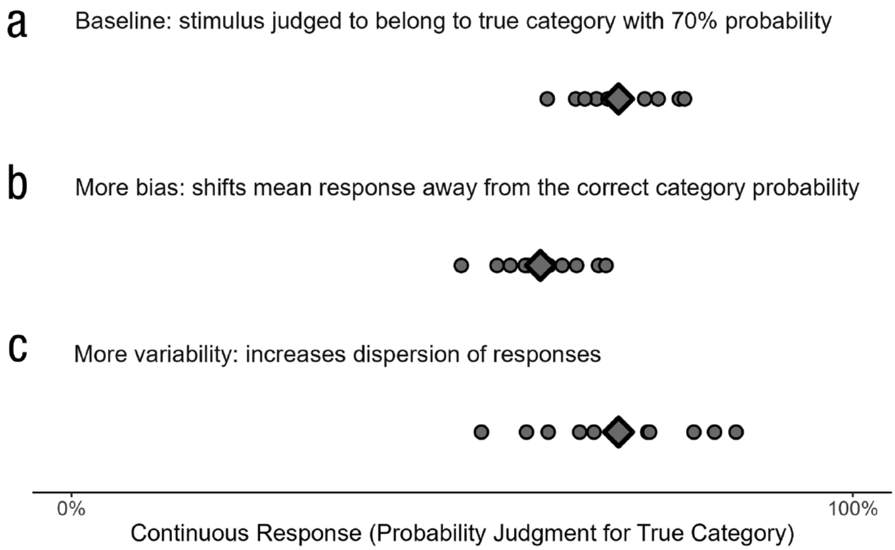

Fig. 7. Independence of bias and variability in continuous data. The points show 10 continuous responses (probability judgments for category membership) of a person to the same object; the diamond is the mean response. More bias increases the difference between the mean response and the true category probability and is thus independent of behavioral variability.

that can check for perceptual noise alongside other more critical stimuli used for model comparison.

Finally and maybe most importantly, it seems probable that at any point in time, multiple forms of noise operate simultaneously. It is unclear how our simulation, which investigated different forms of noise in isolation, would scale up when multiple forms of noise are present at the same time and perhaps even correlated. Future research could look at more complicated combinations of noise and integrate these in a larger simulation encompassing other categorization models and category structures.

Conclusion

Behavioral variability is ubiquitous in human categorization—the same object is sometimes assigned to one category and sometimes to another. In many situations, however, consistent, error-free categorization is preferred, and interventions such as training aim to decrease the variability with which aspiring experts in a domain perform categorizations. Our article aimed to show that looking at behavioral variability can reveal something about the underlying categorization mechanism. Specifically, this article reviewed different sources of behavioral variability in categorizations, focusing on perceptual intake and cognitive processing. In simulations, we showed that it was not possible to determine the source of the variability simply by using category responses. However, the different forms of noise had distinct profiles when analyzing continuous predictions. Assessing people's category beliefs in a continuous way may therefore help disentangle perceptual and process-related sources of behavioral variability. Ultimately, this can inform not only cognitive models but also applied interventions by indicating exactly which stage of the categorization mechanism to target to reduce behavioral variability.

Appendix A

Model simulation results for the six individual category structures

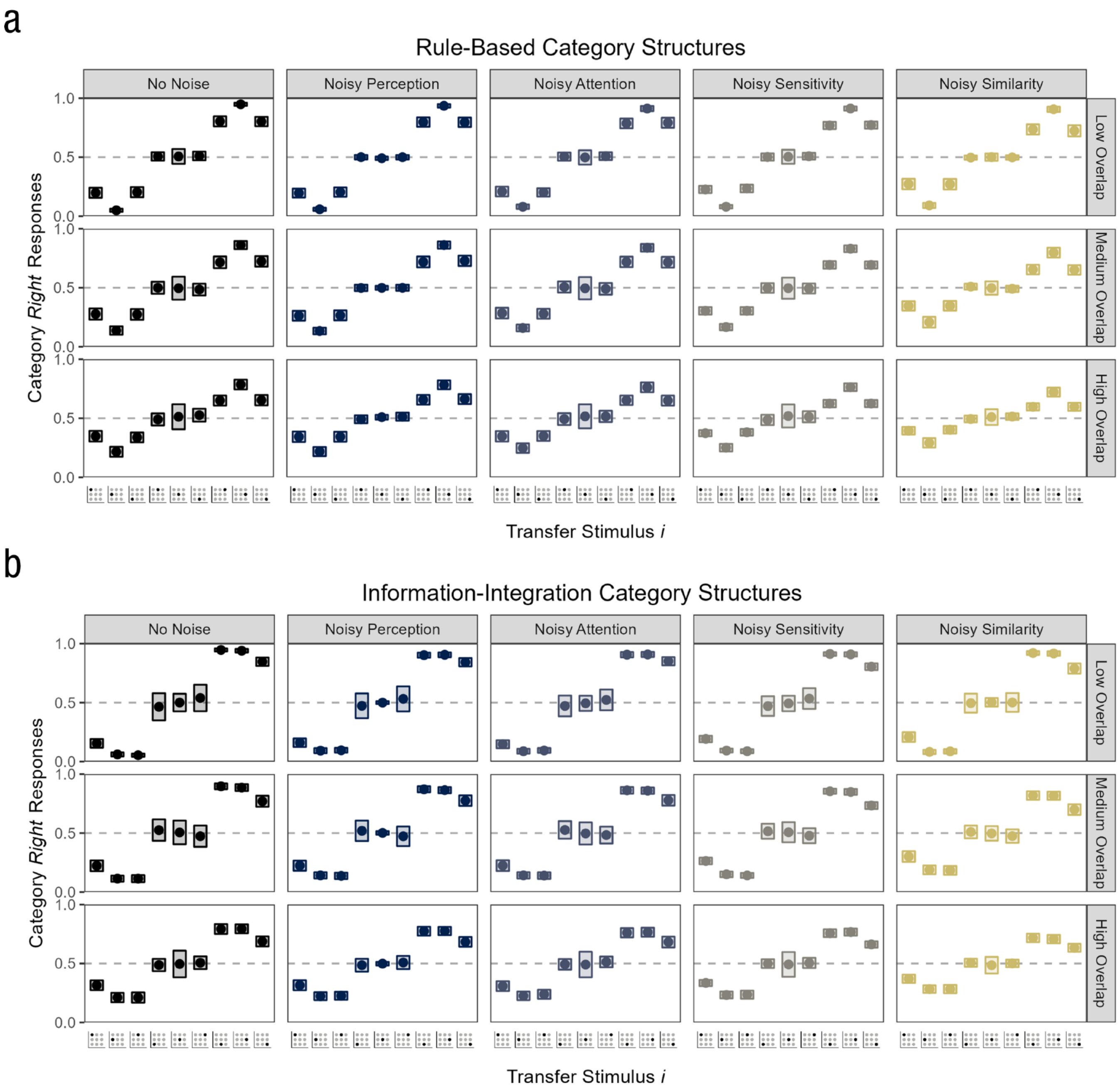

Fig. A1. Simulated binary category responses for all six category structures (for details, see Fig. 5): (a) rule-based category structures; (b) information-integration category structures.

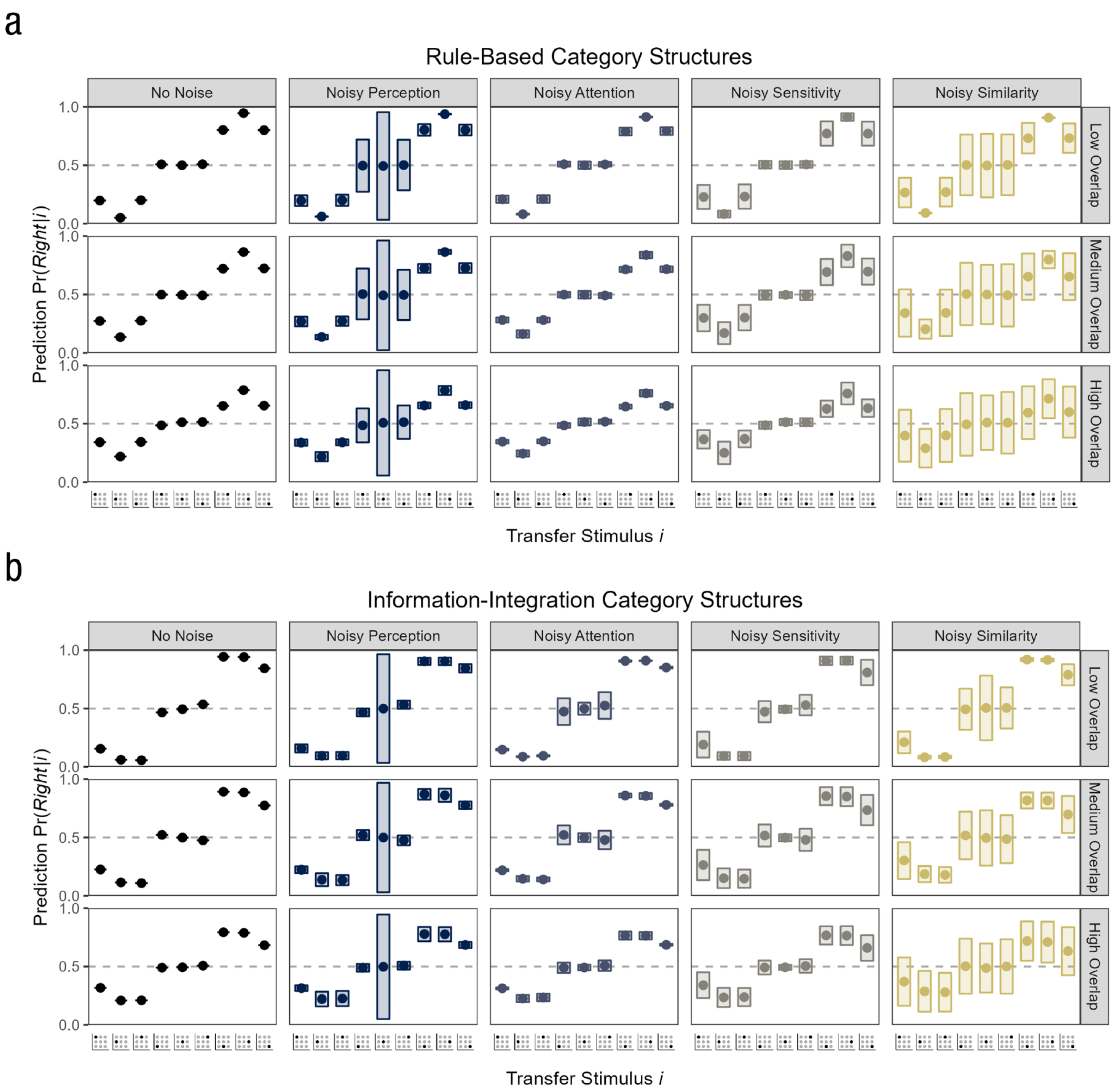

Fig. A2. Simulated continuous category predictions for all six category structures (for details, see Fig. 6): (a) rule-based category structures; (b) information-integration category structures.

Appendix B

Binary parameter recovery: goodness-of-fit measures

In addition to the correlations reported in the main text, we computed the mean absolute error (MAE) and the mean signed difference (MSD) between the true and estimated parameter values. Compared with the MAE, the MSD takes the sign of the difference into account and thus is useful to quantify any bias in the estimated parameter estimates. Specifically, for two vectors with true parameters $\boldsymbol{\theta}$ and estimated parameters $\hat{\boldsymbol{\theta}}$, $MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |\hat{\theta}_i - \theta_i$ and $MSD = \frac{1}{n} \cdot \sum_{i=1}^{n} \hat{\theta}_i - \theta_i$. Figure B1 reports both indices, normalized by each parameter's permissible range of values for comparability, together with the correlations between true and estimated parameter values.

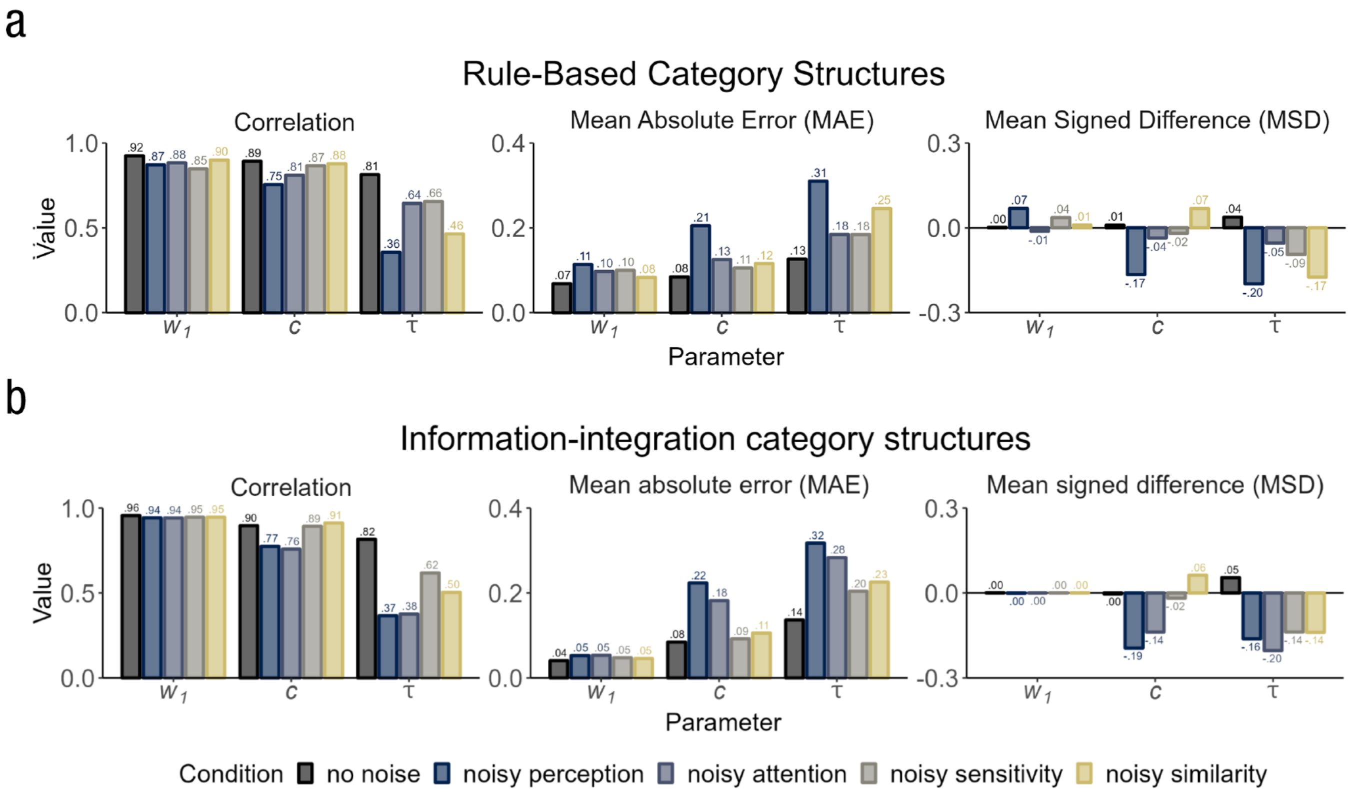

Fig. B1. Parameter recovery indices for the attention weight w_1, the distance sensitivity c, and the response determinism parameter τ under different forms of noise for the (a) rules-based category structures and (b) information-integration structures: the correlation between the true and estimated values, the mean absolute error, and the mean signed difference. The parameter recovery was conducted on simulated binary category responses.

Appendix C

Continuous parameter recovery

In addition to the parameter recovery on the binary category responses reported in the main text, we ran a parameter recovery on continuous responses such as can be obtained in a category probability task. Specifically, for each stimulus i we sampled a continuous response R_i^* from a normal distribution centered around the prediction $\Pr(Right \mid i)$ of Equation 2, $R_i^* \sim \mathcal{N}\left(\Pr(Right \mid i), \sigma\right)$. The standard deviation σ was randomly sampled from a uniform distribution between 0 and .5 (the possible range given responses could range from 0 to 1) and was estimated with the other parameters (the attention weight w, the distance sensitivity c, and the response determinism τ). As in the binary case, the parameter estimation used maximum likelihood and was noise-free—only the simulated responses R_i^* could contain trial-specific noise. Figures C1 and C2 show the parameter recovery results.

The results qualitatively resemble those from the parameter recovery on the binary category responses in two ways. First, the parameters were well recovered in the absence of noise; the mean correlations between the true and estimated values across all six category structures were $r_w = .93$ for attention weight w, $r_c = .87$ for distance sensitivity c, $r_\tau = .73$ for response determinism τ, and $r_\sigma = .99$ for the standard deviation σ. Second, adding noise deteriorated the parameter recovery; the

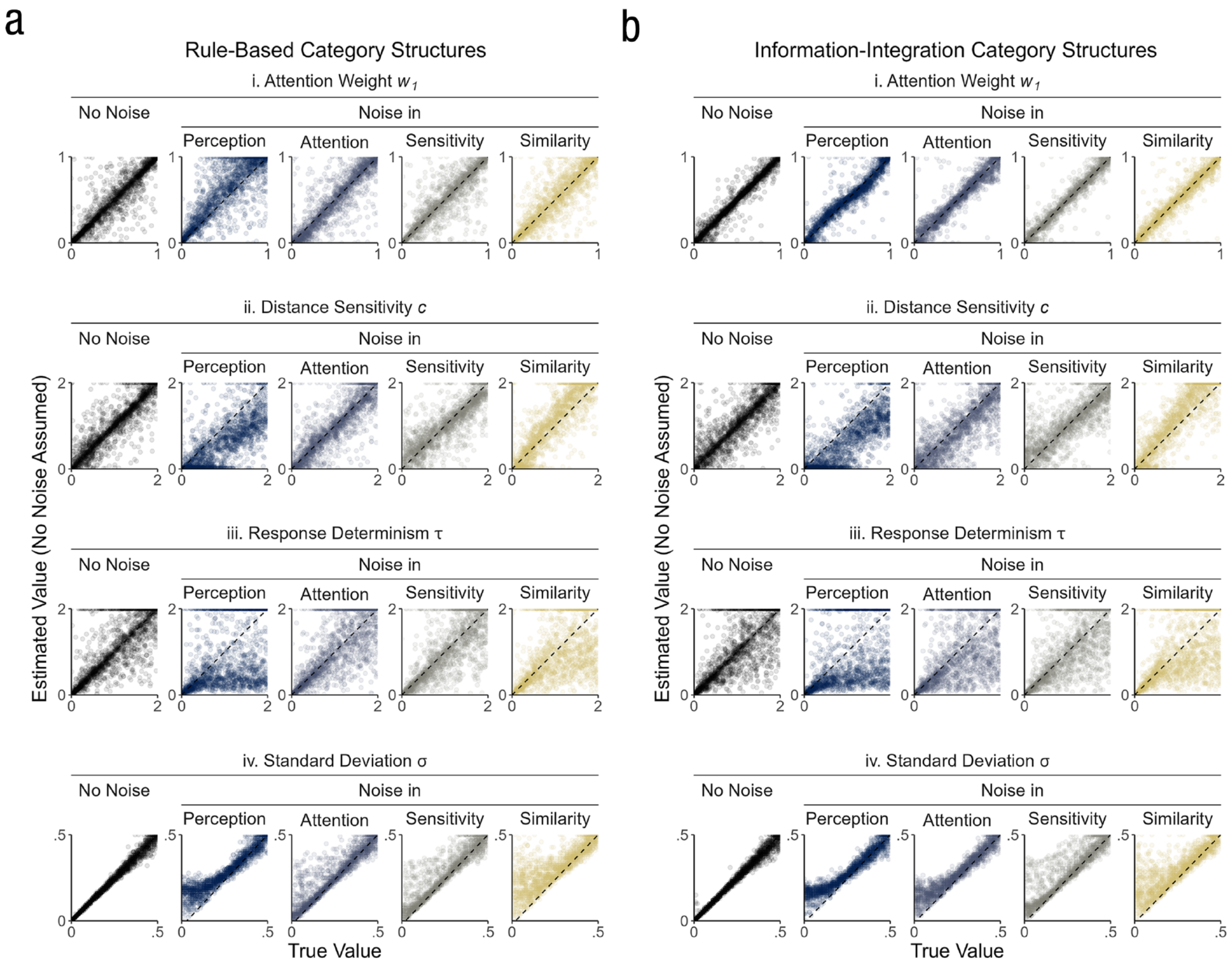

Fig. C1. Continuous parameter recovery for the attention weight w_1, the distance sensitivity c, the response determinism τ, and the standard deviation σ under different forms of noise in the (a) rule-based category structures and (b) information-integration category structures. The x-axis shows the true parameter values, the y-axis shows the estimated values, and the dashed diagonal represents a perfect recovery of the parameters from simulated continuous category responses.

mean correlations across category structures after adding noise were r_w = .90, r_c = .81, r_τ = .57, and r_σ = .90. As in the parameter recovery on the binary category responses, the attention weight w was estimated in a nearly bias-free way (i.e., MSD_w = .01). The distance sensitivity c and the response determinism τ were then underestimated to a lesser extent than in the binary parameter recovery (MSD_c = −.03 and MSD_τ = −.04). The standard deviation σ, on the other hand, was overestimated (MSD_σ = .09). This makes sense because larger values for σ mean more variability in the continuous responses. In other words, σ reflects response determinism in the opposite way as τ: Whereas the small τ tried to capture the variability in the binary data, the large σ did so in the continuous data (see also Seitz, von Helversen, et al., 2023).

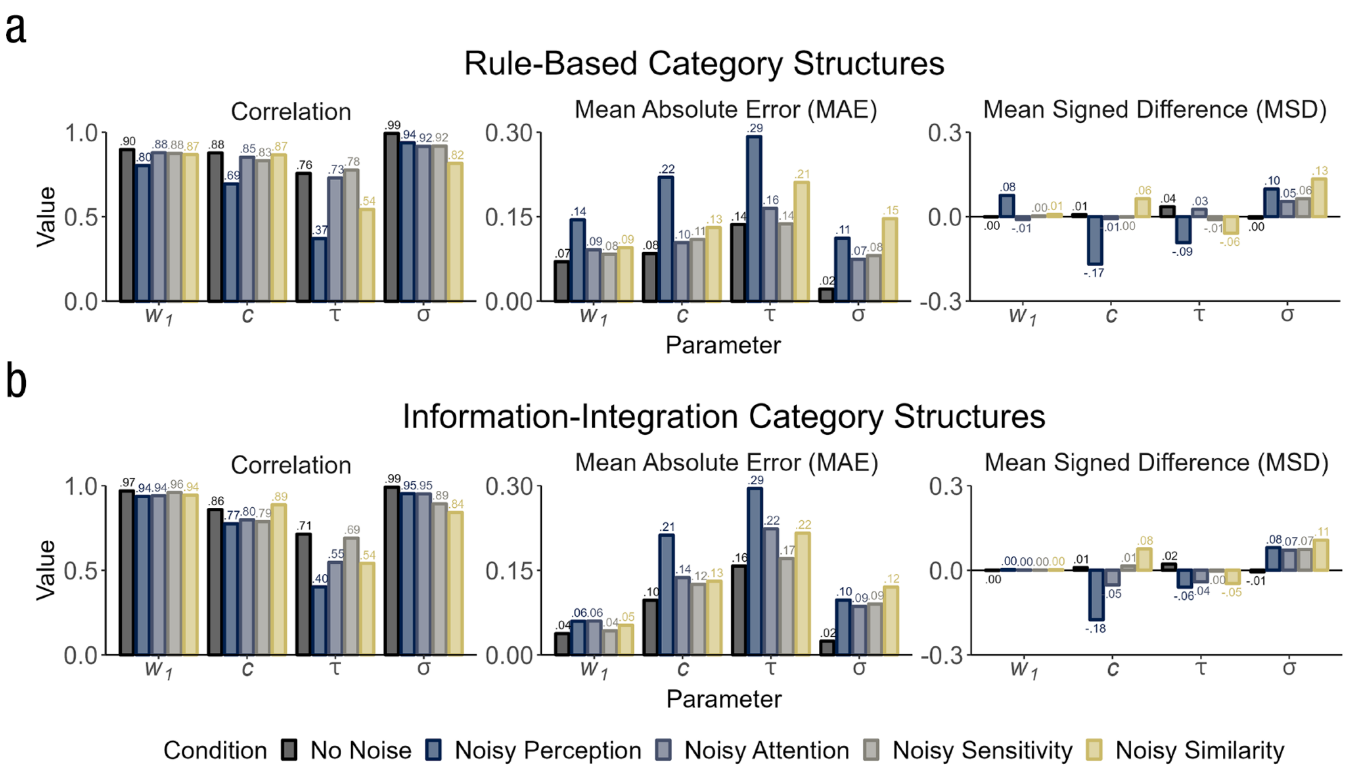

Fig. C2. Parameter recovery indices for the attention weight w_1, the distance sensitivity c, the response determinism parameter τ, and the standard deviation σ under different forms of noise for the (a) rule-based category structures and (b) information-integration category structures: the correlation between the true and estimated values, the mean absolute error, and the mean signed difference. The parameter recovery was conducted on simulated continuous category responses.

Transparency

Action Editor: Joakim Sundh
Editor: Interim Editorial Panel
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This work was supported by Swiss National Science Foundation Grant 195389 (to F. I. Seitz).

ORCID iD

Florian I. Seitz https://orcid.org/0000-0003-3217-4083

Notes

1. Note that this relation between c and the number of exemplars determining the category prediction holds only if the choice rule compares the similarities by their ratio, such as in Luce's (1959) original choice rule.

2. The simulation results are virtually identical if the similarity function uses a city-block distance ($r = 1$).

3. Ell and Ashby used sine-wave gratings as stimuli with the features "spatial frequency" and "orientation."

4. The parameters have a natural lower bound at 0, and the attention weight w_1 cannot exceed 1. The upper bounds for c and τ were set to 2 to ensure that the model made variable predictions for different stimuli (larger values can lead the model to make equally deterministic predictions for all possible stimuli).

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. https://doi.org/10.1037/xlm0000772

Alfonso-Reese, L. A. (2001). Technique for estimating perceptual noise in categorization tasks. *Behavior Research Methods, Instruments, & Computers*, *33*(4), 489–495. https://doi.org/10.3758/BF03195407

Alfonso-Reese, L. A., Ashby, F. G., & Brainard, D. H. (2002). What makes a categorization task difficult? *Perception & Psychophysics*, *64*, 570–583. https://doi.org/10.3758/bf03194727

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology:*

General, *120*(1), 3–19. https://doi.org/10.1037/0096-3445.120.1.3

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review*, *8*(4), 629–647. https://doi.org/10.3758/BF03196200

Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Psychology Press. https://doi.org/10.4324/9781315807607

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, *44*(2), 310–329. https://doi.org/10.1006/jmps.1998.1249

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. https://doi.org/10.1037/0033-295x.105.3.442

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*(5), 204–210. https://doi.org/10.1016/S1364-6613(00)01624-7

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), 33–53. https://doi.org/10.1037//0278-7393.14.1.33

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

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 598–612. https://doi.org/10.1037/0096-1523.16.3.598

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 50–71. https://doi.org/10.1037/0096-1523.18.1.50

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*(3), 372–400. https://doi.org/10.1006/jmps.1993.1023

Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement, judgment and decision making* (pp. 251–301). Elsevier. https://doi.org/10.1016/B978-012099975-0.50006-3

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*(1), 124–150. https://doi.org/10.1037/0033-295X.95.1.124

Ashby, F. G., Smith, J. D., & Rosedahl, L. A. (2020). Dissociations between rule-based and information-integration categorization are not caused by differences in task difficulty. *Memory & Cognition*, *48*, 541–552. https://doi.org/10.3758/s13421-019-00988-4

Balakrishnan, J., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 615–633. https://doi.org/10.1037//0096-1523.22.3.615

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*, Article 5418. https://doi.org/10.1038/s41467-020-18946-z

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: The role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30–39. https://doi.org/10.1016/j.neuron.2012.03.016

Bröder, A., Gräf, M., & Kieslich, P. J. (2017). Measuring the relative contributions of rule-based and exemplar-based processes in judgment: Validation of a simple model. *Judgment and Decision Making*, *12*(5), 491–506. https://doi.org/10.1017/S1930297500006513

Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.

Cavagnaro, D. R., & Regenwetter, M. (2023). Probabilistic choice induced by strength of preference. *Computational Brain & Behavior*, *6*(4), 569–600. https://doi.org/10.1007/s42113-023-00176-3

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

Collsiöö, A., Sundh, J., & Juslin, P. (2023). Unpacking intuitive and analytic memory sampling in multiple-cue judgment. In K. Fiedler, P. Juslin, & J. Denrell (Eds.), *Sampling in judgment and decision making* (pp. 177–204). Cambridge University Press. https://doi.org/10.1017/9781009002042.010

Edmunds, C., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *Quarterly Journal of Experimental Psychology*, *68*(6), 1203–1222. https://doi.org/10.1080/17470218.2014.978875

Ell, S. W., & Ashby, F. G. (2006). The effects of category overlap on information-integration and rule-based category learning. *Perception & Psychophysics*, *68*(6), 1013–1026. https://doi.org/10.3758/BF03193362

Ennis, D. M., & Johnson, N. L. (1993). Thurstone-Shepard similarity models as special cases of moment generating functions. *Journal of Mathematical Psychology*, *37*(1), 104–110. https://doi.org/10.1006/jmps.1993.1005

Ennis, D. M., Palen, J. J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology*, *32*(4), 449–465. https://doi.org/10.1016/0022-2496(88)90023-5

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107–140. https://doi.org/10.1037/0096-3445.127.2.107

Erickson, M. A., & Kruschke, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, *9*(1), 160–168. https://doi.org/10.3758/BF03196273

Estes, Z. (2004). Confidence and gradedness in semantic categorization: Definitely somewhat artifactual, maybe absolutely natural. *Psychonomic Bulletin & Review*, *11*(6), 1041–1047. https://doi.org/10.3758/BF03196734

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, *109*(3), 416–422. https://doi.org/10.1016/j.cognition.2008.09.007

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585–612. https://doi.org/10.1146/annurev.psych.49.1.585

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43. https://doi.org/10.1016/s0010-0277(00)00099-8

Goldstone, R. L., & Son, J. Y. (2012). Similarity. In K. J. Holyoak & M. R. G (Eds.), *Oxford library of psychology: The Oxford handbook of thinking and reasoning* (pp. 155–176). Oxford University Press.

Hahn, U., Prat-Sala, M., Pothos, E. M., & Brumby, D. P. (2010). Exemplar similarity and rule application. *Cognition*, *114*(1), 1–18. https://doi.org/10.1016/j.cognition.2009.08.011

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. https://doi.org/10.1038/s41562-020-00951-3

Herzog, S. M., and von Helversen, B. (2018). Strategy selection versus strategy blending: A predictive perspective on single- and multi-strategy accounts in multiple-cue estimation. *Journal of Behavioral Decision Making*, *31*, 233–249. doi:10.1002/bdm.1958.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation's blindsight: How cognitive load can improve judgments. *Psychological Science*, *24*(6), 869–879. https://doi.org/10.1177/0956797612463581

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, *143*(6), 2242–2261. https://doi.org/10.1037/a0037989

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2016). Similar task features shape judgment and categorization processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1193–1217. https://doi.org/10.1037/xlm0000241

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 418–439. https://doi.org/10.1037/0278-7393.7.6.418.

Izydorczyk, D., & Bröder, A. (2023). Measuring the mixture of rule-based and exemplar-based processes in judgment: A hierarchical Bayesian approach. *Decision*, *10*(4), 347–371. https://doi.org/10.1037/dec0000195

Jarecki, J. B., Meder, B., & Nelson, J. D. (2018). Naïve and robust: Class-conditional independence in human classification learning. *Cognitive Science*, *42*(1), 4–42. https://doi.org/10.1111/cogs.12496

Jarecki, J. B., & Seitz, F. I. (2020). Cognitivemodels: An R package for formal cognitive modeling. In T. C. Stewart (Ed.), *Proceedings of the 18th International Conference on Cognitive Modelling* (pp. 100–106). Applied Cognitive Science Lab, Penn State.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 316–332. https://doi.org/10.1037/0278-7393.32.3.316

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*(1), 259–298. https://doi.org/10.1016/j.cognition.2007.02.003

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133–156. https://doi.org/10.1037/0096-3445.132.1.133

Karlsson, L., Juslin, P., & Olsson, H. (2007). Adaptive changes between cue abstraction and exemplar memory in a multiple-cue judgment task with continuous cues. *Psychonomic Bulletin & Review*, *14*(6), 1140–1146. https://doi.org/10.3758/bf03193103

Krantz, D. H., & Tversky, A. (1975). Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology*, *12*(1), 4–34. https://doi.org/10.1016/0022-2496(75)90047-4

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

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 267–301). Cambridge University Press. https://doi.org/10.1017/CBO9780511816772

Lacroix, G. L., Giguere, G., & Larochelle, S. (2005). The origin of exemplar effects in rule-driven categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 272–288. https://doi.org/10.1037/0278-7393.31.2.272

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, *124*(2), 161–180. https://doi.org/10.1037/0096-3445.124.2.161

Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 695–711. https://doi.org/10.1037/0278-7393.24.3.695

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227–260. https://doi.org/10.3758/BF03206876

Lamberts, K., & Brockdorff, N. (1997). Fast categorization of stimuli with multivalued dimensions. *Memory & Cognition*, *25*(3), 296–304. https://doi.org/10.3758/BF03211285

Lee, J. C., Lovibond, P. F., Hayes, B. K., & Navarro, D. J. (2019). Negative evidence and inductive reasoning in generalization of associative learning. *Journal of Experimental Psychology: General*, *148*(2), 289–303. https://doi.org/10.1037/xge0000496

Lee, M., & Wetzels, R. (2010). Individual differences in attention during category learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*(32), 387–392.

Little, D. R., & Lewandowsky, S. (2009a). Better learning with more error: Probabilistic feedback increases sensitivity to correlated cues in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1041–1061. https://doi.org/10.1037/a0015902

Little, D. R., & Lewandowsky, S. (2009b). Beyond nonutilization: Irrelevant cues can gate learning in probabilistic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 530–550. https://doi.org/10.1037/0096-1523.35.2.530

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Lovibond, P. F., Lee, J. C., & Hayes, B. K. (2020). Stimulus discriminability and induction as independent components of generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(6), 1106–1120. https://doi.org/10.1037/xlm0000779

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

Maddox, W. T. (2001). Separating perceptual processes from decisional processes in identification and categorization. *Perception & Psychophysics*, *63*(7), 1183–1200. https://doi.org/10.3758/bf03194533

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70. https://doi.org/10.3758/BF03211715

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 650–662. https://doi.org/10.1037/0278-7393.29.4.650

Maddox, W. T., & Bogdanov, S. V. (2000). On the relation between decision rules and perceptual representation in multidimensional perceptual categorization. *Perception & Psychophysics*, *62*(5), 984–997. https://doi.org/10.3758/bf03212083

Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1459–1482. https://doi.org/10.1037//0278-7393.24.6.1459

Maddox, W. T., & Bohil, C. J. (2004). Probability matching, accuracy maximization, and a test of the optimal classifier's independence assumption in perceptual categorization. *Perception & Psychophysics*, *66*(1), 104–118. https://doi.org/10.3758/BF03194865

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & David, A. (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. https://doi.org/10.1037/0278-7393.30.1.227

Mata, R., von Helversen, B., Karlsson, L., & Cüpper, L. (2012). Adult age differences in categorization and multiple-cue judgment. *Developmental Psychology*, *48*(4), 1188–1201. https://doi.org/10.1037/a0026084

Meagher, B. J., & Nosofsky, R. M. (2023). Testing formal cognitive models of classification and old-new recognition in a real-world high-dimensional category domain. *Cognitive Psychology*, *145*, Article 101596. https://doi.org/10.1016/j.cogpsych.2023.101596

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

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 275–292. https://doi.org/10.1037/0278-7393.28.2.275

Minda, J. P., & Smith, J. D. (2011). Prototype models of categorization: Basic formulation, predictions, and limitations. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 40–64). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.003

Navarro, D. J., & Lee, M. D. (2002). Combining dimensions and features in similarity-based representations. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15 (NIPS 2002)*. MIT Press. https://proceedings.neurips.cc/paper_files/paper/2002/file/243facb29564e7b448834a7c9d901201-Paper.pdf

Newell, B. R., & Bröder, A. (2008). Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making*, *3*(3), 195–204. https://doi.org/10.1017/S1930297500002400

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. H. Ross (Ed.), *Psychology of learning and motivation* (pp. 167–215). Elsevier.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104–114. https://doi.org/10.1037/0278-7393.10.1.104

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 87–108. https://doi.org/10.1037/0278-7393.13.1.87

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics*, *45*(4), 279–290. https://doi.org/10.3758/BF03204942

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* (pp. 149–167). Lawrence Erlbaum Associates.

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 282–304. https://doi.org/10.1037//0278-7393.15.2.282

Nosofsky, R. M., & Hu, M. (2023). Category structure and region-specific selective attention. *Memory & Cognition*, *51*(4), 915–929. https://doi.org/10.3758/s13421-022-01365-4

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

Nosofsky, R. M., & Little, D. R. (2010). Classification response times in probabilistic rule-based category structures: Contrasting exemplar-retrieval and decision-boundary models. *Memory & Cognition*, *38*(7), 916–927. https://doi.org/10.3758/MC.38.7.916

Nosofsky, R. M., Meagher, B. J., & Kumar, P. (2022). Contrasting exemplar and prototype models in a natural-science category domain. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(12), 1970–1994. https://doi.org/10.1037/xlm0001069

Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review*, *3*(2), 222–226. https://doi.org/10.3758/BF03212422

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/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. https://doi.org/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. https://doi.org/10.1037/0033-295X.101.1.53

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018). Tests of an exemplar-memory model of classification learning in a high-dimensional natural-science category domain. *Journal of Experimental Psychology: General*, *147*(3), 328–353. https://doi.org/10.1037/xge0000369

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

Olschewski, S., & Rieskamp, J. (2021). Distinguishing three effects of time pressure on risk taking: Choice consistency, risk preference, and strategy selection. *Journal of Behavioral Decision Making*, *34*(4), 541–554. https://doi.org/10.1002/bdm.2228

Olschewski, S., Rieskamp, J., & Scheibehenne, B. (2018). Taxing cognitive capacities reduces choice consistency rather than preference: A model-based test. *Journal of Experimental Psychology: General*, *147*(4), 462–484. https://doi.org/10.1037/xge0000403

Palmeri, T. J., Wong, A. C., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Sciences*, *8*(8), 378–386. https://doi.org/10.1016/j.tics.2004.06.001

Peterson, J. C., Battleday, R. M., Griffths, T. L., & Russakovsky, O. (2019). Human uncertainty makes classification more robust. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9617–9626). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICCV.2019.00971

Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *Journal of Neuroscience*, *31*(47), 17220–17229. https://doi.org/10.1523/JNEUROSCI.2028-11.2011

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285–293. https://doi.org/10.1016/j.tics.2015.03.002

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences*, *28*(1), 1–14. https://doi.org/10.1017/s0140525x05000014

Richler, J. J., & Palmeri, T. J. (2014). Visual category learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 75–94. https://doi.org/10.1002/wcs.1268

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631–661. https://doi.org/10.1257/jel.44.3.631

Roads, B. D., & Love, B. C. (2023). Modeling similarity and psychological space. *Annual Review of Psychology*, *75*, 215–240. https://doi.org/10.1146/annurev-psych-040323-115131

Rodrigues, P. M., & Murre, J. M. (2007). Rules-plus-exception tasks: A problem for exemplar models? *Psychonomic Bulletin & Review*, *14*(4), 640–646. https://doi.org/10.3758/BF03196814

Rosner, A., Schaffner, M., & von Helversen, B. (2022). When the eyes have it and when not: How multiple sources of activation combine to guide eye movements during multiattribute decision making. *Journal of Experimental Psychology: General*, *151*(6), 1394–1418. https://doi.org/10.1037/xge0000833

Rouder, J. N., & Ratcliff, R. (2006). Comparing exemplar- and rule-based theories of categorization. *Current Directions in Psychological Science*, *15*(1), 9–13. https://doi.org/10.1111/j.0963-7214.2006.00397.x

Sakamoto, Y., Matsuka, T., & Love, B. C. (2004). Dimension-wide vs. exemplar-specific attention in category learning and recognition. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the 6th International Conference on Cognitive Modeling* (pp. 261–266). Lawrence Erlbaum Associates.

Sanders, C. A., & Nosofsky, R. M. (2020). Training deep networks to construct a psychological feature space for a natural-object category domain. *Computational Brain & Behavior*, *3*(3), 229–251. https://doi.org/10.1007/s42113-020-00073-z

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*, *120*(1), 39–64. https://doi.org/10.1037/a0030777

Scheibehenne, B., von Helversen, B., & Rieskamp, J. (2015). Different strategies for evaluating consumer products: Attribute-and exemplar-based approaches compared. *Journal of Economic Psychology*, *46*, 39–50. https://doi.org/10.1016/j.joep.2014.11.006

Schlegelmilch, R., & von Helversen, B. (2020). The influence of reward magnitude on stimulus memory and stimulus generalization in categorization decisions. *Journal of Experimental Psychology: General*, *149*(10), 1823–1854. https://doi.org/10.1037/xge0000747

Schlegelmilch, R., Wills, A. J., & von Helversen, B. (2021). A cognitive category-learning model of rule abstraction, attention learning, and contextual modulation. *Psychological Review*, *129*(6), 1211–1248. https://doi.org/10.1037/rev0000321

Scholz, A., von Helversen, B., & Rieskamp, J. (2015). Eye movements reveal memory processes during similarity- and rule-based decision making. *Cognition*, *136*, 228–246. https://doi.org/10.1016/j.cognition.2014.11.019

Seitz, F. I. (2023). Relative attention across features predicts that common features increase geometric similarity. In *Proceedings of the 21st Annual Meeting of the International Conference on Cognitive Modelling* (pp. 217 –222). Applied Cognitive Science Lab, Penn State.

Seitz, F. I., Albrecht, R., von Helversen, B., Rieskamp, J., & Rosner, A. (2024). *Identifying similarity- and rule-based processes in quantitative judgments: A multi-method approach combining cognitive modeling and eye tracking* [Manuscript submitted for publication]. Center for Economic Psychology, Department of Psychology, University of Basel.

Seitz, F. I., Jarecki, J. B., & Rieskamp, J. (2023). *Perceptual similarity mostly ignores within-category feature distributions: Evidence from computational modeling of human categorizations*. PsyArXiv. https://doi.org/10.31234/osf.io/7v95h

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*, Article 105358. https://doi.org/10.1016/j.cognition.2022.105358

Serre, T. (2016). Models of visual categorization. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(3), 197–213. https://doi.org/10.1002/wcs.1385

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*(2), 125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, *27*(3), 219–246.

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

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42. https://doi.org/10.1037/h0093825

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. https://doi.org/10.1037//0096-3445.121.3.278

Sieck, W. R., & Yates, J. F. (2001). Overconfidence effects in category learning: A comparison of connectionist and exemplar memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 1003–1021. https://doi.org/10.1037/0278-7393.27.4.1003

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), 1411–1436. https://doi.org/10.1037/0278-7393.24.6.1411

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3–27. https://doi.org/10.1037/0278-7393.26.1.3

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 800–811. https://doi.org/10.1037/0278-7393.28.4.800

Stevens, S. S. (1960). The psychophysics of sensory function. *American Scientist*, *48*(2), 226–253.

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. https://doi.org/10.1037//0278-7393.28.1.3

Sundh, J., Collsiöö, A., Millroth, P., & Juslin, P. (2021). Precise/not precise (PNP): A Brunswikian model that uses judgment error distributions to identify cognitive processes. *Psychonomic Bulletin & Review*, *28*, 351–373. https://doi.org/10.3758/s13423-020-01805-9

Thibaut, J.-P., & Gelaes, S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1403–1415. https://doi.org/10.1037/0278-7393.32.6.1403

Thibaut, J.-P., Gelaes, S., & Murphy, G. L. (2018). Does practice in category learning increase rule use or exemplar use—or both? *Memory & Cognition*, *46*, 530–543. https://doi.org/10.3758/s13421-017-0782-4

Trippas, D., & Pachur, T. (2019). Nothing compares: Unraveling learning task effects in judgment and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(12), 2239–2266. https://doi.org/10.1037/xlm0000696

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Vanpaemel, W. (2009). BayesGCM: Software for Bayesian inference with the generalized context model. *Behavior Research Methods*, *41*(4), 1111–1120. https://doi.org/10.3758/BRM.41.4.1111

Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin & Review*, *15*, 732–749. https://doi.org/10.3758/PBR.15.4.732

Verbeemen, T., Vanpaemel, W., Pattyn, S., Storms, G., & Verguts, T. (2007). Beyond exemplars and prototypes as memory representations of natural concepts: A clustering approach. *Journal of Memory and Language*, *56*(4), 537–554. https://doi.org/10.1016/j.jml.2006.09.006

Verguts, T., & Fias, W. (2009). Similarity and rules united: Similarity-and rule-based processing in a single neural network. *Cognitive Science*, *33*(2), 243–259. https://doi.org/10.1111/j.1551-6709.2009.01011.x

von Helversen, B., Herzog, S. M., & Rieskamp, J. (2014). Haunted by a doppelgänger: Irrelevant facial similarity affects rule-based judgments. *Experimental Psychology*, *61*(1), 12–22. https://doi.org/10.1027/1618-3169/a000221

von Helversen, B., Karlsson, L., Mata, R., & Wilke, A. (2013). Why does cue polarity information provide benefits in inference problems? The role of strategy selection and knowledge of cue importance. *Acta Psychologica*, *144*(1), 73–82. https://doi.org/10.1016/j.actpsy.2013.05.007

von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue abstraction processes in multiple-cue judgment. *Developmental Psychology*, *46*(1), 220–229. https://doi.org/10.1037/a0016690

Wills, A. J., Inkster, A. B., & Milton, F. (2015). Combination or differentiation? Two theories of processing order in classification. *Cognitive Psychology*, *80*, 1–33. https://doi.org/10.1016/j.cogpsych.2015.04.002

Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin*, *138*(1), 102–125. https://doi.org/10.1037/a0025715

Wyart, V., & Koechlin, E. (2016). Choice variability and suboptimality in uncertain environments. *Current Opinion in Behavioral Sciences*, *11*, 109–115. https://doi.org/10.1016/j.cobeha.2016.07.003

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

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1160–1173. https://doi.org/10.1037/0278-7393.29.6.1160