PSYKHE
2022, 31(SI 1), 1-15
https://doi.org/10.7764/psykhe.2021.37971

ISSN 0717-0297
www.psykhe.cl

Computational Cognitive models of Categorization: Predictions under Conditions of Classification Uncertainty

Modelos Cognitivos Computacionales de Categorización: Predicciones bajo Condiciones de Clasificación Incierta

Nicolás Marchant[1] and Sergio E. Chaigneau[1, 2]

[1] Centro de Neurociencia Social y Cognitiva, Escuela de Psicología, Universidad Adolfo Ibáñez
[2] Centro de Investigación de la Cognición, Escuela de Psicología, Universidad Adolfo Ibáñez

In the category learning literature, similarity models have monopolized a good deal of research. The prototype and exemplar models are both based on the idea that people represent the structure of categories and category instances in the physical world in a mental similarity space. However, evidence for these models comes mainly from paradigms that provide subjects with deterministic feedback (i.e., exemplars belong to their corresponding categories with probability = 1). There is evidence that results obtained with deterministic feedback paradigms may not generalize well under probabilistic feedback conditions (i.e., where exemplars belong to their corresponding categories with probability less than 1). In this current work, we also suggest that probabilistic feedback may better reflect natural conditions, which is another important reason to pursue probabilistic feedback research. Thus, in the current work we set up a category learning experiment with probabilistic feedback and use it to evaluate different models. In addition to the two similarity models discussed above, we also use an associationist model that does not rely on the similarity construct. To compare our three models, we rely on computational modeling, which is a standard way of model comparison in cognitive psychology. Our results show that our associationist model outperforms similarity models on all our model evaluation measures. After presenting our results, we discuss why the similarity-based models fail, and also suggest some future lines of research that are possible using probabilistic feedback procedures.

Keywords: category learning, categorization, probabilistic learning, computational cognitive models

Los modelos basados en similitud han dominado la literatura de aprendizaje de categorías. El modelo de prototipos y el de ejemplares se basan en la idea de que las personas representan los objetos en el mundo real en un espacio de similitud subjetiva. Sin embargo, la evidencia a favor de estos modelos viene predominantemente de paradigmas que usan retroalimentación determinista (i.e., los ejemplares pertenecen a sus categorías correspondientes con una probabilidad igual a 1). Hay evidencia de que los resultados obtenidos con retroalimentación determinista podrían no generalizarse bajo condiciones de retroalimentación probabilista (i.e., cuando los ejemplares pertenecen a sus respectivas categorías con probabilidad menor a 1). En el presente trabajo, también sugerimos que la retroalimentación probabilística podría reflejar mejor las condiciones naturales, lo que es otra buena razón para desarrollar investigación con retroalimentación probabilística. En consecuencia, en el presente trabajo implementamos un experimento de aprendizaje de categorías con retroalimentación probabilística y lo usamos para evaluar distintos modelos. Además de los dos modelos de similitud discutidos más arriba, usamos también un modelo asociacionista que no depende del constructo de similitud. Para comparar nuestros modelos, usamos modelamiento computacional, que es un modo estándar para comparar modelos en psicología cognitiva. Nuestros resultados muestran que el modelo asociacionista supera a los modelos basados en similitud en todas las medidas de evaluación. Luego de presentar nuestros resultados, discutimos por qué fallan los modelos de similitud y también sugerimos líneas futuras de investigación que son posibles al usar procedimientos con retroalimentación probabilista.

Palabras clave: aprendizaje de categorías, categorización, aprendizaje probabilístico, modelos cognitivos computacionales

Nicolás Marchant https://orcid.org/0000-0001-7891-1005

Sergio E. Chaigneau https://orcid.org/0000-0001-8642-6325

Este estudio recibió apoyo económico de la Universidad Adolfo Ibáñez a través una beca de doctorado al primer autor, y de la Agencia Nacional de Investigación y Desarrollo (ANID) a través del proyecto FONDECYT Regular 1190006 otorgado al segundo autor. El artículo es parte de la tesis para Optar al Grado de Doctor en Neurociencia Social y Cognición de la Universidad Adolfo Ibáñez. No existe ningún conflicto de intereses que revelar.

La correspondencia relativa a este artículo debe ser dirigida a Nicolás Marchant, Centro de Neurociencia Social y Cognitiva, Escuela de Psicología, Universidad Adolfo Ibáñez, Avda. Presidente Errázuriz 3328, Las Condes, Santiago, Región Metropolitana, Chile. Email: nicolasmarchant@alumnos.uai.cl

Categorization is a crucial cognitive ability that has drawn attention of psychologists concerned with how the human mind classifies different classes of objects or events into subsequent categories. This ability seems vital for human survival and adaptation to the environment (Seger & Miller, 2010). Although categorization is a broad cognitive ability which impacts many different aspects of the everyday life (e.g., deciding if a pop-up window is a virus or an add), we focus here on the specific process of how people acquire categories. This sub-field within categorization research has been referred to as category learning. In most of category learning studies (Ashby & Maddox, 2005; 2011), researchers focus on how people learn, represent and select different categories. In those studies, a typical procedure is that subjects receive extensive classification training and are then tested on how well they learned the trained categories (i.e., generalization of learning).

However, most of category learning experiments have relied on deterministic feedback (i.e., the feedback signal assumes no uncertainty relative to how exemplars are classified) (Marchant & Chaigneau, 2021). The literature focusing on how people acquire categories under conditions of uncertainty is sparse (Knowlton et al., 1994; Lagnado et al., 2006; Little & Lewandowsky, 2009a; 2009b; Meeter et al., 2008). Here, we deep dive on why categorization under conditions of uncertainty might be important to study. To that end, we implemented a novel category learning experiment which allowed us to directly manipulate classification probabilities. Furthermore, we tested two famous categorization formal models (i.e., the similarity-based models) alongside with a novel model of associative learning. By reporting these novel experimental results, we expect to contribute to elucidating how people acquire categories in the context of uncertainty.

Categorization under Uncertainty

To address the issue of classification under uncertainty, we implemented a category learning experiment that trained subjects by giving them probabilistic feedback. But, why use probabilistic routines on training? We can think of at least three reasons for why designing category learning tasks with probabilistic feedback may be important: ecological validity, task dependency of experimental results, and differences in strategies and brain mechanisms that underlie those strategies. We discuss them next.

As mentioned above, the most common type of feedback in category learning experiments is deterministic feedback (DF; Ashby & Ell, 2001; Nosofsky et al., 1994). In DF, each feature combination (i.e., exemplar) is always a member of one of the categories. For example, in a DF environment, features on an exemplar like “has four legs”, “barks” and “has fur” is always a member of the “dog” category. Consequently, subjects’ performance increases during training towards some asymptotic performance level (i.e., perfect or close to perfect performance). In contrast, probabilistic feedback (PF) has been used much less frequently (Ashby & Gott, 1988; Gluck & Bower, 1988; Little & Lewandowsky, 2009a; 2009b; and in the Weather Prediction Task, Knowlton et al., 1994; Meeter et al., 2008). In PF, each exemplar (i.e., feature combination) belongs to a category with a probability less than 1.0. Following our previous example, an exemplar that “barks”, “has four legs” and “has fur” is member of the “dog” category only on 80 % of the trials. In contrast to DF, training under PF conditions continue to provide corrective feedback even when subjects achieve asymptotic performance, which never approaches perfect performance. Elsewhere (Marchant & Chaigneau, 2021), we have argued that PF should be used more frequently, given that probabilistic conditions may be more representative of natural learning environments where different sources of feedback may offer inconsistent feedback (e.g., think of being taught the same subject by different teachers; Little & Lewandowsky, 2009a; Meeter, et al., 2008; Lagnado, et al., 2006) and that at least some empirical findings in category learning may be conditional on using deterministic feedback. For example, inter-feature correlations have for a long time been considered an important part of conceptual representations (Hoffman & Rehder, 2010; Ell et al., 2017). It is generally accepted that subjects in category learning experiments tend not to learn inter-feature correlations. For inter-feature correlations to be learned, inference tasks have to be used (Chin-Parker & Ross, 2002; Yamauchi et al., 2002). However, when probabilistic feedback has been used with classification procedures, evidence has been found that subjects do learn inter-feature correlations (Little & Lewandowsky, 2009a). Consequently, it is possible that categorization phenomena being uncovered by using DF fail to generalize under PF conditions. Given these concerns, PF should be used more broadly.

Another reason for using PF is that it seems to induce a different strategy and brain mechanisms than DF. It has been shown that traditional category learning procedures with DF allow subjects to easily elaborate a form of logical declarative rule through explicit reasoning and hypothesis testing (Ashby et al., 1998; Ashby & Valentin, 2017). Because feedback is consistently linked to the desired response, subjects can discover the logical rule that allows them to categorize and generalize accurately (e.g., “if feature x is present,

then the exemplar belongs to category A"; Ashby & Valentine, 2018). In contrast, when experimental scenarios increase in uncertainty (i.e., when the desired response is probabilistic), people seem to rely less on a rule-based declarative strategy. This is precisely what has been found in research using the Weather Prediction Task (WPT) with amnesic patients (Knowlton et al., 1994; Gluck et al., 1996). The WPT is a probabilistic feedback task in which subjects receive cues that are probabilistically associated with the outcome (i.e., rain or shine), and have to learn to probabilistically use them to predict the weather. In that study, amnesic patients performed normally relative to a control group in a first set of 50 training trials but performed significantly worse than controls during the subsequent trials. The explanation for this pattern of results is that, while controls were able to find the rule that allowed categorization by trial 50 and to use it during the subsequent trials, amnesic patients were not able to hold in memory an explicit rule and therefore failed relative to controls. In contrast, patients with Basal Ganglia dysfunction such as Parkinson's and Huntington's disease showed an early impairment during the WPT (Knowlton et al., 1996; Knowlton et al., 1996). Because the Basal Ganglia are widely thought to implement procedural learning (Lawrance et al., 1998; Shohamy et al., 2008; Seger & Miller, 2010), this whole pattern of results is consistent with the idea that probabilistic feedback promotes a procedural route to categorization, in contrast to deterministic feedback, which promotes a declarative or explicit route.

Computational Models of Categorization

Like in many other areas of cognitive psychology (e.g., memory, attention, learning, reasoning and categorization), researchers in category learning have implemented computational models (Kruschke, 2008; Richler & Palmeri, 2014). Computational models of cognition have emerged as a response to the necessity of a mathematical formalization that represents the functionality of the cognitive system. By using these, researchers have been able to make predictions, to contrast different models, and to modify them to improve their explanatory power (e.g., Lewandowsky & Farrell, 2011).

As discussed in Wills and Pothos (2012), having computational or mathematical implementations affords advantages. By comparing competing models' ability to fit the empirical data, rigorous comparisons can be made. By formalizing theories, hypotheses can be unambiguously formulated (i.e., in contrast to purely verbal theories). Because models are sometimes complex, they produce behavioral predictions that are not evident at first sight, thus providing deeper insights into empirical phenomena. In what follows next, we discuss three models of category learning, together with their mathematical and algorithmic implementations, all of which allows the model fitting results we report in the current work. Because similarity-based models (discussed next) have been dominant in the literature for the past 50 years, in our work we wanted to test their performance in PF conditions, and to contrast it with an associationist model. In the next section, we discuss the models, beginning with similarity-based models of category learning.

Similarity-Based Models

Most of the currently dominant categorization models rely on the similarity assumption. The construct of similarity has a long and important tradition in cognitive psychology, traceable at least to the Gestalt movement. In short, similarity assumes that when two items are similar, the distance between them in a hypothetical psychological space is narrowed. In contrast, when both items are dissimilar, the psychological space between them is expanded (Shepard, 1987; Nosofsky, 1984; 1986) (e.g., a "Golden Retriever" is more similar to a "Beagle" than any of them is to a "Coyote").

Interestingly for us, Kruschke (2008) points out that categorization models require at least three specifications: (1) the representation of internal category knowledge, (2) the process of matching a to-be-classified item with the subjective representation of the category and (3) a process of choice-response over a category. As it will become clear next, the matching process between item and category in the following models is a similarity-based process.

Exemplar Model

Exemplar models assume that people's category representations are specific memory-traces of individual items belonging to that category. Consequently, when subjects are asked to categorize an item x into category A, the theory assumes that a comparison between item x and all previous experienced items is computed through a similarity-based process.

The most prominent computational model of the exemplar theory is Nosofsky's Generalized Context Model (GCM, Nosofsky, 1984) which is a generalization of Medin and Schaffer's Context Model (1978). The GCM assumes that selective attention guides the similarity-based process, meaning that an *x* item's most attended to (or relevant) attributes are the ones which contribute the most to the perceived similarity. Additionally, similarity is bound to change through the learning period due to the continuous optimization of the attentional resources (Nosofsky, 2011).

The discussed above ideas allow formalizing the distance between an item *x* to exemplar *y* as shown in the following equation:

$$d(x,y) = \sum_{i=1}^{m} (w_i |x_i - y_i|^r)^{1/r} \tag{1}$$

Where $d(x,y)$ denotes the psychological distance, x_i represents the value of item *x* on the i-th dimension (or attribute) of its *m* binary dimensions ($m = 3$ in the following experiment), and y_i represents the individual training exemplar on its i-th binary dimensions. The attentional parameter is represented by w_i, meaning that attention can be allocated to each of stimulus dimension (with the constraint that there is a fixed total amount of attention to be allocated; in our computational modeling, this means that the total sum of all attentional coefficients sums to 1). The original version of the GCM uses an exponential parameter *r* which is set equal to 1 (city-block metric) when perceptual separable stimuli are used in the experiment, which is the case of this study. Consequently, the similarity of item *x* to category A is formalized as:

$$S_A(x) = \sum_{y \in A} \exp\left(-c\, d(x,y)\right) \tag{2}$$

Where $S_A(x)$denotes similarity. The sensitivity scale parameter *c* corresponds to an exponential decaying function of distance. Thus, when the distance is 0, then the similarity is 1. Finally, to obtain the predicted probability of the category A given each item *x*, Luce's choice axiom:

$$p(A|x) = S_A(x)^\gamma / S_A(x)^\gamma + S_B(x)^\gamma \tag{3}$$

Where, the similarity to category A is divided by the summed similarity to category A and B, thus bounding p(A | x) to the 0 to 1 range. Note, that Ashby & Maddox (1993) suggested a γ parameter which controls the deterministic course of responses. This means that when $\gamma > 1$ subjects tend to respond more deterministically, on the contrary, when $\gamma < 1$ subjects tend to respond more probabilistically. Gamma achieves this by essentially warping the equation's output, moving it from a flat straight line to progressively closer to a step function.

Prototype Model

Prototype models assume that people create an abstract prototypical representation of the category; a kind of central tendency of all previous experienced item attributes. When subjects are asked to categorize an item *x* into category A, the theory assumes that people compare item *x* with the prototype of category A by means of a similarity-based process.

Nosofsky and Zaki (2002) elaborated a Multiplicative Prototype Model (MPM) which enables direct comparison with the GCM. The MPM assumes a multiplicative response choice equal to the GCM. Formally:

$$S_A(x) = \exp\left[-c \sum_{i=1}^{m} (w_i |x_i - proto_{Ai}|^r)^{1/r}\right] \tag{4}$$

Where $proto_{Ai}$ is the prototype of category A. This formulation integrates in the same equation the distance metric, which is similar to eq. 1 and the similarity **metrics** which is similar to eq. 2. Note that to obtain p(category | item), MPM uses the same choice axiom showed in eq. 3, including a γ parameter.

Associative Learning

Adaptive Linear Filter

The Adaptive Filter model (Widrow & Hoff, 1960) views category learning as a task where two alternative responses (A or B category) are reinforced based on how features are combined. Note that this is consistent with the idea that the Basal Ganglia are involved in procedural category learning, as previously discussed (Ashby & Ennis, 2006). Rather than creating a similarity space where instances are represented, the Adaptive Filter model assumes that people update coefficients for each feature (f_1, f_2, f_3, …, f_k), such that classification errors relative to feedback are minimized. The aforementioned coefficients are more formally the Least Mean Square (LMS) correlation coefficients relating each feature with the classification criterion that is operative during an experiment. Importantly, the model assumes that learning occurs only when errors are made (i.e., when corrective feedback is received; Rescorla & Wagner, 1972; Schultz, 1998; Little & Lewandowsky, 2009a). Widrow and Hoff (1960; see also Widrow & Kamenetsky, 2003) show that the algorithm in eq. (5) converges to the LMS correlation coefficients relating each feature with the classification criterion. Gluck and Bower (1988) tested this aspect of the theory and reported data consistent with people's error correcting mechanism being able to converge toward a LMS solution. Eq. (5) below shows the LMS learning algorithm.

$$w_{j+i}^{i} = w_{j}^{i} + n(d - r)x^{i} \tag{5}$$

Where, $w_{(j+1)}$ is the adjusted weight for feature fi due to the performance in the previous trial, w_j is the weight for that same feature in the immediately preceding trial, n is a learning rate, d is the desired response for the preceding trial, r is the actual response provided by the subject for the exemplar received in the preceding trial, and x is the state of feature f_i during the preceding trial.

As a result of learning, subjects' responses will be a function of a linear combination of features states and their corresponding weights, as shown in the parenthetical term in eq. (6). That linear equation is deterministic. As discussed in Gluck and Bower (1988), a simple model that relates coefficients to categorization probabilities is the logistic probability function (eq. (6)).

$$p(A) = 1/[1 + e^{-g(\beta_0+\beta_1 f_1+\beta_2 f_2+\beta_3 f_3+\cdots+\beta_k f_k)}] \tag{6}$$

Where the betas are the aforementioned correlation coefficients, the f's are each of the features that describe exemplars, which may be in state either 1 or -1, and g is a parameter that allows representing different sensitivities with which the linear term is transformed into categorization probability (i.e., the slope of the sigmoid function). Note that g in eq. (6) behaves similarly to the γ parameter in eq. (3).

Method

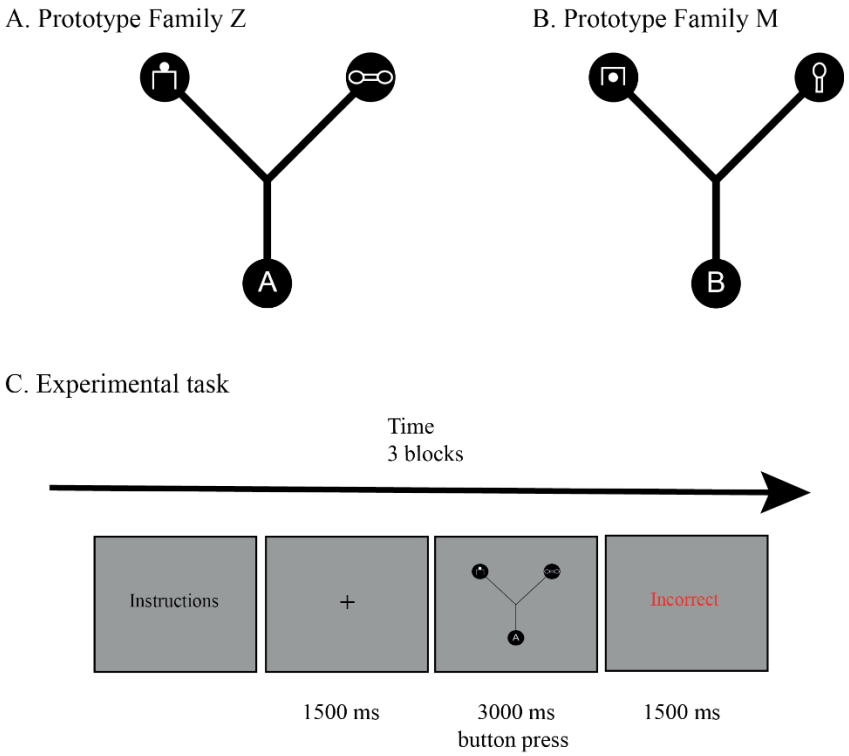
In the experiment, we report how participants learned to classify eight exemplars constructed by combining three features (f_1, f_2, f_3) that could adopt one of two possible states. Importantly, though feedback was provided, it did not imply a consistent exemplar to category classification (i.e., probabilistic feedback was used). Because we were interested in the learning process (i.e., how people acquire categories), we used artificial categories with novel features to prevent participants' prior knowledge from impacting results.

Design and Materials

The design was a 3 (feature association strength: $f_1 > f_2 > f_3$, $f_2 > f_3 > f_1$, $f_3 > f_1 > f_2$) x 3 (features: f_1, f_2, f_3) mixed design, with the last being the repeated measures factor. Stimuli were constructed such that the association between each feature and the classification criterion followed the order shown above. Also, stimuli were constructed as to reflect all possible combinations of the 3 features (f_1, f_2, f_3) that could assume one of two possible states (i.e., "ceremonial symbols", see Figure 1; similar to those used in Rehder et al., 2009). With these 3 features, there are 8 possible feature combinations or exemplars ($2^3 = 8$; see Table 1). By design, only two features contributed to classification in our experiment (i.e., the third feature was always irrelevant), and an exemplar was never completely associated with one of the categories (i.e., neither A nor B). Exemplars 1 and 2 in Table 1 belonged to category A with p(A) = .9 (i.e., they belonged to category B with p(B) = 1 - p(A) = .1). Exemplars 3 and 4 belonged to category A with p(A) = .7 (i.e., they belonged

to category B with p(B) = 1 - p(A) = .3). The category boundary was symmetric, such that for exemplars 5 and 6 p(B) = .7 (and p(A) = .3), and for exemplars 7 and 8 p(B) = .9 and p(A) = .1) (see Table 2). Importantly, these probabilities defined the way in which participants would receive corrective PF (e.g., p(A) = .9 means that a subject that consistently responded A to the corresponding exemplar, would receive a "correct" feedback on 90 % of those trials, but "incorrect" feedback on 10 % of those trials). Exemplars constructed as described above resulted in features that show different degrees of association with the classification criterion (i.e., r = .6, r = .2, r = 0; see Table 3).

Figure 1
Complete Experiment Set-Up



Note. (A) Prototype ceremonial symbol for Family Z. (B) Prototype ceremonial symbol for Family M. Note that other exemplars are created by distorting these two prototypes. (C) Experimental procedure consisting of 240 trials.

Table 1
Every Exemplar is Shown as an Effect Coded Combination of Binary Valued Features (f_1, f_2 and f_3)

Exemplar	*f1*	*f2*	*f3*
E1	1	1	1
E2	1	1	-1
E3	1	-1	1
E4	1	-1	-1
E5	-1	1	1
E6	-1	1	-1
E7	-1	-1	1
E8	-1	-1	-1

Note. Features are uncorrelated to each other.

Table 2
Classification Probability to Category A for each Condition by every Exemplar

Exemplar	Condition 1 $f_1>f_2>f_3$ p(A)	Condition 2 $f_2>f_3>f_1$ p(A)	Condition 3 $f_3>f_1>f_2$ p(A)
E1	0.9	0.9	0.9
E2	0.9	0.7	0.3
E3	0.7	0.3	0.9
E4	0.7	0.1	0.3
E5	0.3	0.9	0.7
E6	0.3	0.7	0.1
E7	0.1	0.3	0.7
E8	0.1	0.1	0.1

Note. Classification probability to category B is 1 - p(A). These probabilities can be used to guide how probabilistic feedback is provided (see text for details).

Table 3
Individual Feature Association Weights (i.e., r_{xy}) for each Condition

Feature	Condition 1 $f_1>f_2>f_3$	Condition 2 $f_2>f_3>f_1$	Condition 3 $f_3>f_1>f_2$
f_1	0.6	0.0	0.2
f_2	0.2	0.6	0.0
f_3	0.0	0.2	0.6

Participants and Procedures

Thirty-six undergraduate students (27 females) aged 18 to 37 (mean = 20.11, *SD* = 3.21) signed informed consent to participate in the experiment for course credit. Sample size was estimated based on previous experiments. However, results suggest that our sample size choice was appropriate (Cohen's $F = 0.53$). The informed consent was made in accordance with the Adolfo Ibáñez University Ethics Committee. Participants were randomly assigned to one of the three between subjects' experimental conditions: Condition 1 ($f_1 > f_2 > f_3$), Condition 2 ($f_2 > f_3 > f_1$) and Condition 3 ($f_3 > f_1 > f_2$); twelve participants in each condition). The experiment lasted approximately 30 minutes.

In our experiment, participants learned to classify 8 exemplars into one of two categories (A or B). During training, participants were provided with trial-by-trial corrective PF with the schedule shown in Table 2. Participants underwent 3 blocks of 80 trials each, for a total of 240 training trials. During each block, participants received 10 repetitions of the 8 exemplars (i.e., 10 x 8 = 80). Trial order was randomized within each block. Subjects had to press the keyboard button "Z" if they believed the presented exemplar belonged into category A, otherwise, they had to press the button "M". For correct responses, a green "correct" sign appeared on screen for 1500 ms. For incorrect responses, a red "incorrect" sign appeared on the screen for the same amount of time. Subjects had 30 seconds to give a response, otherwise a "too slow" message appeared on the screen (see Figure 1).

As shown in Figure 1, three spheres covering 9.62 cm^2 each are the ceremonial symbols on the screen. The spheres remained in the same screen location during the complete experiment. The experiment was built using PsychoPy v3 and mounted online through the Pavlovia environment (Peirce et al., 2019). We used the last block (i.e., block 3) to perform the model fits we present in the Modeling Results section.

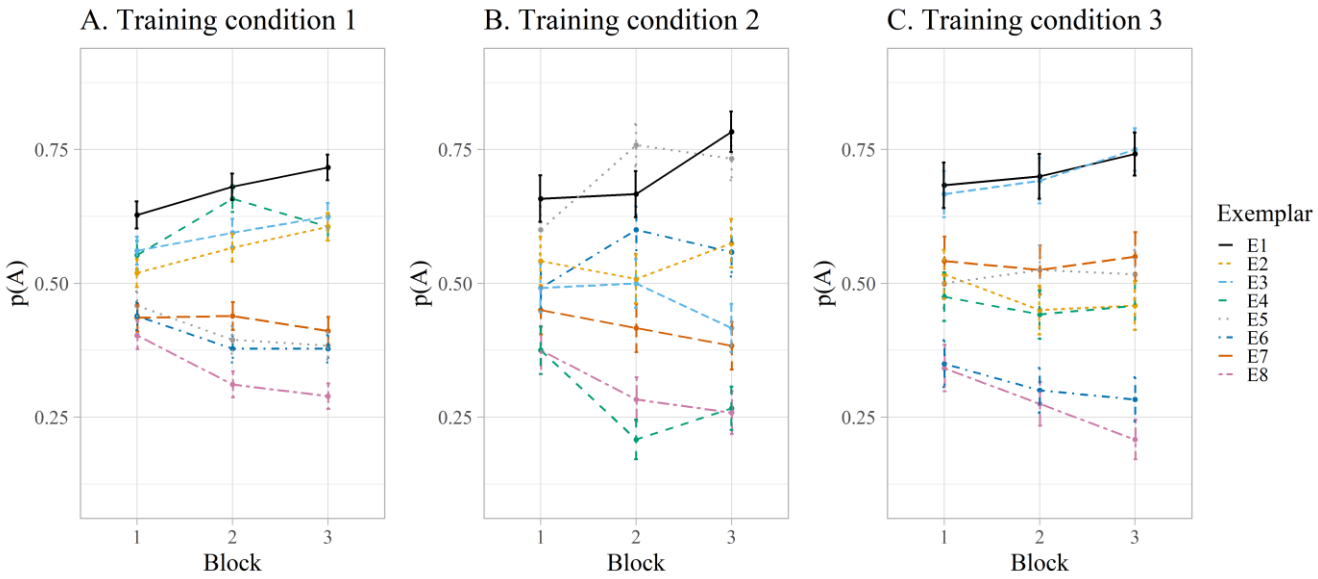
Results

Behavioral Results

Behavioral results revealed that participants in condition 1 ($f_1 > f_2 > f_3$) achieved a mean accuracy of .58 (SD = .06), in condition 2 ($f_2 > f_3 > f_1$) mean accuracy was .63 (SD = .08), and in condition 3 ($f_3 > f_1 > f_2$) mean accuracy was .62 (SD = .07). In none of the conditions did subjects approach optimal performance, which was set at .77 (calculated by Luce's axiom). Furthermore, it is apparent that none of the subjects showed signs of using only the most diagnostic rule for classification. If subjects only had used the most diagnostic feature in each condition, they would approach the 77 % correct response criterion. We clearly did not find this (see Figure 2).

A factorial 3 (conditions: $f_1 > f_2 > f_3$, $f_2 > f_3 > f_1$ and $f_3 > f_1 > f_2$) x 3 (blocks) design with the last being the repeated measure factor, revealed a main effect of block ($F(2,66) = 9.55$, MSe = .05, $p < .001$, $\eta_p^2 = .22$, power = .98), a non-significant main effect of condition (p = .21), and a non-significant interaction between block and condition (p = .83). These results suggest that there was a learning effect across blocks in every condition. Contrast comparisons revealed a significant difference between block 3 and the previous blocks (blocks 1 and 2) ($F(1,33) = 9.02$, MSe = .07, p = .005, $\eta_p^2 = .22$, power = .83), and a significant difference between blocks 1 and 2 ($F(1,33) = 9.99$, MSe = .12, p = .003, $\eta_p^2 = .23$, power = .87). This comparison shows that by block 3 subjects had learned to discriminate between exemplars (see Figure 2).

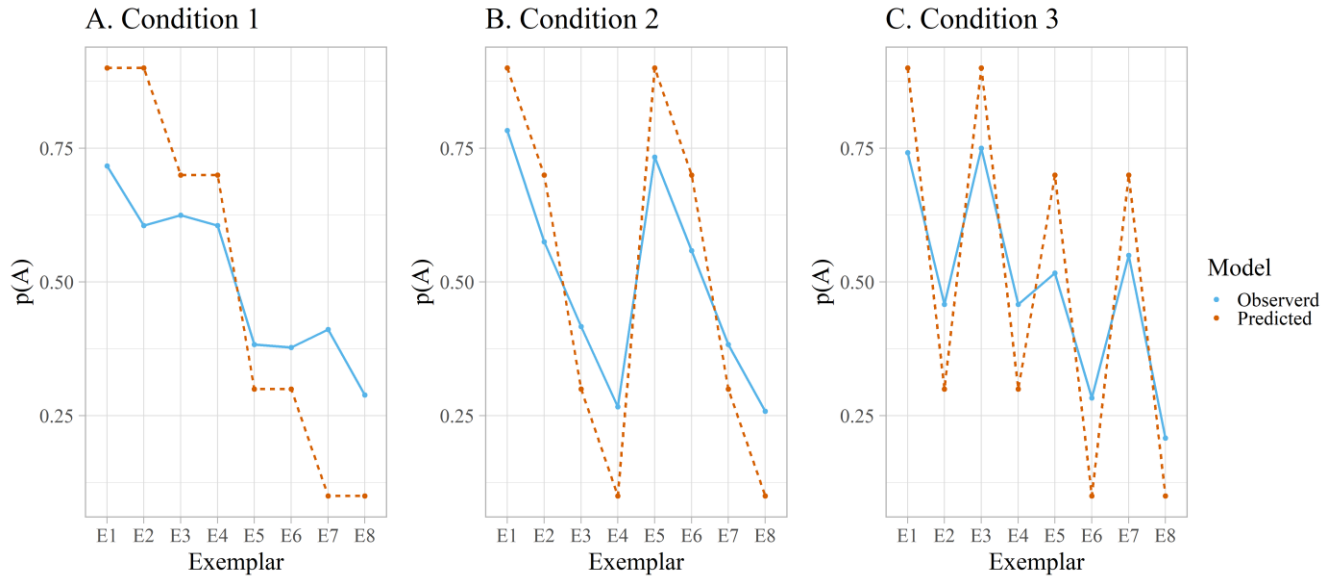
Figure 2
Classification Probability of Category A (Family Z) given each Specific Exemplar Combination across Blocks



Additionally, we compared our classification predictions using the Adaptive Filter with the empirical training accuracy in block 3. The predictions of the Adaptive Filter are given by eq. (6). For these predictions, we averaged accuracy in block 3 for each exemplar and in each condition (recall that contrast comparisons showed that in block 3 had achieved asymptotic performance). Because empirical probabilities and predicted probabilities are two linear trajectories, we directly estimated the R^2 between both. As Figure 3 shows, there is a clear tendency for subjects in every condition to learn the exemplar classification probability. In general, fittings are fairly good. For condition 1 ($f_1 > f_2 > f_3$), R^2 = .90; for condition 2 ($f_2 > f_3 > f_1$), R^2 = .97; and for condition 3 ($f_3 > f_1 > f_2$), R^2 = .89.

Figure 3

Probability of Classification in Category A (button Z) for a given Exemplar in each Condition



Note. (A) Condition 1 ($f_1 > f_2 > f_3$), (B) Condition 2 ($f_2 > f_3 > f_1$) and (C) condition 3 ($f_3 > f_1 > f_2$). Continuous lines show average subject data on block 3 for each exemplar, and the dashed line shows optimal probabilities by experimental task presented in Table 2. Note, that the optimal classification probabilities are computed by Eq. (6).

These comparisons revealed that by providing PF subjects were able to learn to classify close to what the Adaptive Filter stipulates. In other words, by simply contrasting trajectories we could observe that the Adaptive Filter accounts for human classification data using PF. However, there are still gaps regarding the cognitive mechanism that underlie probabilistic classification. For this issue, a more formal comparison is advised. In the next section, we contrast the Adaptive Filter predictions against similarity-based models using a model fitting strategy.

Modeling Results

GCM, MPM and ALF were fitted for each subject using p(A|exemplar) computed from frequencies obtained during training block 3. As shown earlier, training results suggest that there was a learning effect across blocks in every condition. Thus, we assume that during block 3 subjects had already learned to accurately discriminate between exemplars (at least as accurately as they were able to). Consequently, for each subject and model we computed the predicted p(A|exemplar) probability for block 3 and adjusted each separate model's free parameters using the negative log-likelihood (lnL, see eq. 7; Little & Lewandowsky, 2009a) as an error metric, computed as:

$$lnL = -\sum_i d_i \ln(p_i) + (n_i - d_i)\ln(1 - p_i) \tag{7}$$

Where d_i is the observed number of A responses made for exemplar i, p_i is the model predicted probability of category A for exemplar i and n_i is the number of times exemplar i was presented (in this case $n_i = 10$). Because we used negative lnL, lower lnL values indicate a better model fit. In general, closer predicted to observed probabilities indicate a better model. However, parameter interpretability should also be considered when judging model fit.

Both GCM and MPM have five free parameters: a sensitivity scaling parameter (c), three attentional weights (w, fixed to sum 1) and a scaling response parameter ($gamma$). For ALF, we estimated a single free parameter: the scaling response parameter (g). We used standard maximum likelihood methods for parameter estimation with the "fminsearchbnd" function in MatLab (D'Errico, 2021). The parameters were optimized separately for each model and for each subject.

For model comparison we used BIC (Bayesian Information Criterion) and AIC (Akaike Information Criterion). Both are *lnL* (log Likelihood) computations that penalize models with more parameters (in general, BIC performs a stronger penalization than AIC).

Our modeling results show that ALF **achieves** better fits for each of the individual subjects in every condition. In Table 4, fit values averaged across participants are presented. The reader will note that ALF shows better fits, even without penalizing the number of free parameters (with the only exception of condition 3, where ALF is almost tied with the prototype model). Model comparison is even more favorable to ALF when the number of free parameters is penalized (see Table 4).

Table 4
Mean Grouped Fit Values for each Condition and for each Model (GCM, MPM and ALF)

Model	Condition 1			Condition 2			Condition 3		
	lnL	BIC	AIC	lnL	BIC	AIC	lnL	BIC	AIC
GCM	53.52 (1.92)	118.56 (3.85)	117.04 (3.85)	50.97 (3.46)	113.45 (6.92)	111.94 (6.92)	51.84 (3.38)	115.20 (6.75)	113.68 (6.75)
MPM	61.50 (30.73)	134.52 (61.47)	133.0 (61.47)	50.07 (6.86)	111.65 (13.72)	110.14 (13.72)	48.51 (4.69)	108.52 (9.39)	107.01 (9.39)
ALF	52.45 (3.16)	107.19 (6.32)	106.89 (6.32)	47.83 (5.14)	97.97 (10.27)	97.67 (10.27)	48.90 (5.39)	100.09 (10.79)	99.79 (10.79)

Note. Mean values were obtained by averaging subjects' fits for each condition and each model. In parenthesis the standard deviation.

Models' parameter analysis

Additionally, we show the averaged parameter estimations for each of the free parameters for each model in Table 5. Note that values presented are averages across fits obtained for individual subjects and not the result of fits performed on averaged data (for a discussion of these different approaches, see Ashby et al., 1994). For the interested reader, individual level fits are available at https://osf.io/eackz/

Table 5
Mean Parameter Estimates for each Model and each Condition

Model	Condition 1					Condition 2					Condition 3				
	g	*c*	*w1*	*w2*	*w3*	*g*	*c*	*w1*	*w2*	*w3*	*g*	*c*	*w1*	*w2*	*w3*
GCM	0.33 (0.24)	1.06 (0.11)	0.30 (0.02)	0.36 (0.02)	0.34 (0.01)	0.53 (0.30)	1.03 (0.07)	0.34 (0.01)	0.31 (0.02)	0.35 (0.02)	0.46 (0.31)	1.02 (0.03)	0.34 (0.01)	0.35 (0.02)	0.31 (0.01)
MPM	0.17 (0.24)	9.81 (15.12)	0.61 (0.31)	0.28 (0.28)	0.11 (0.15)	0.26 (0.17)	4.77 (4.97)	0.22 (0.35)	0.51 (0.28)	0.27 (0.14)	0.27 (0.21)	2.21 (1.54)	0.42 (0.14)	0.09 (0.17)	0.50 (0.23)
ALF	0.74 (0.56)	-	-	-	-	1.35 (0.75)	-	-	-	-	1.23 (0.75)	-	-	-	-

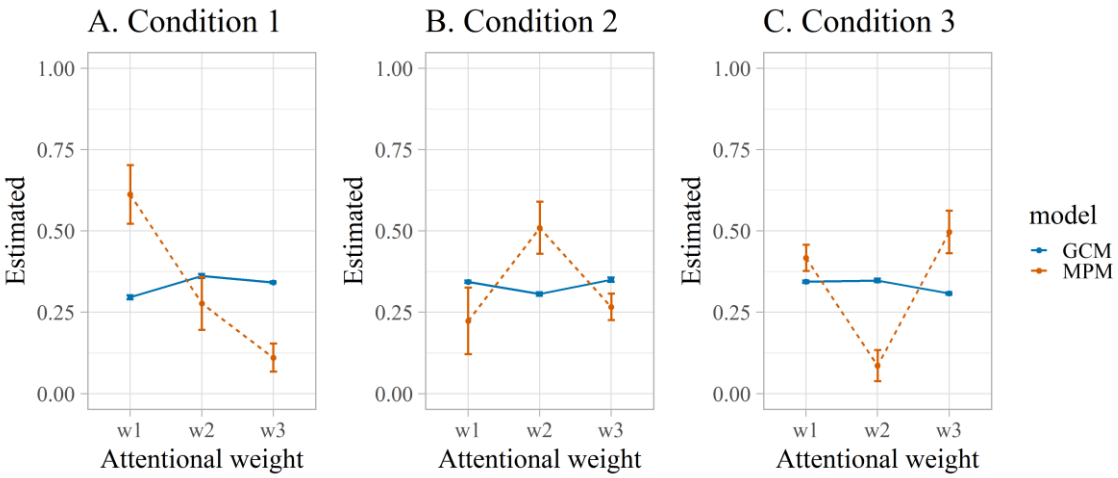
Note. Standard deviation in parenthesis.

An interesting result to look at is the attentional weights parameter *w*. A model that treats *w* as free parameters should aim at capturing the attentional weight we defined in our experiment (see Table 3). Our modeling results show that the GCM cannot capture the feature-relevance pattern specified in Table 3. As shown in Table 5, the GCM estimates the same weights for all three features. In contrast, the MPM shows a sensitivity of the *w* parameters in accordance with feature-weights stipulated during training in each condition (see Table 3). For example, as Figure 4 shows, in condition 1 ($f_1 > f_2 > f_3$), the MPM estimates a higher value of *w* for feature 1, and lower values for features 2 and 3, consistently with weight shown in Table 3. The same occurs for conditions 2 ($f_2 > f_3 > f_1$) and 3 ($f_3 > f_1 > f_2$). This difference in *w* parameter

estimation between GCM and MPM was confirmed through a 3 (condition) x 2 (model: GCM and MPM) x 3 (attentional weights) repeated ANOVA, with the last being the repeated measure factor. For this analysis, we reordered attentional weight data to be coherent to the condition's correlations shown in Table 3. For example, for condition $f_2 > f_3 > f_1$, the second attentional weight (i.e., f_2) is the most important, the third (i.e., f_3) is less important and the first (i.e., f_1) is non-diagnostic. Thus, attentional weights have been ranked. Having reordered data for the three conditions, the repeated measures ANOVA using Greenhouse-Geisser correction revealed a main effect of attentional weight rank ($F(2,132) = 13.26$, $MSe = .66$, $p < .001$, $\eta_p^2 = .17$, power = .99) a significant interaction between attentional weight rank and model ($F(2,132) = 20.35$, $MSe = .99$, $p < .001$, $\eta_p^2 = .24$, power = .99), a non-significant interaction between attentional weight rank and condition ($p = .40$) and a non-significant three way interaction ($p = .28$). This analysis confirmed that MPM attentional weights are statistically different to those estimated by the GCM (i.e., the two-way significant interaction between attentional weight rank and model). Visual inspection of Figure 4 clearly illustrates where the difference lies. Under PF conditions, the GCM cannot capture the purported attentional weights we designed our experiment with (see Table 3), settling on about equal weights for all features (i.e., the almost flat lines in Figure 4). This is an important result because it poses a clear limit to the GCM's explanatory scope.

Figure 4

Attentional Weights Estimated by both Models GCM and MPM and for each Condition



Note. (A) Condition 1 ($f_1 > f_2 > f_3$), (B) Condition 2 ($f_2 > f_3 > f_1$) and (C) Condition 3 ($f_3 > f_1 > f_2$). Solid line show GCM predicted weights and dashed line show MPM predicted weights. Standard errors at 95 %.

Finally, an important aspect of similarity-based models is the *gamma* parameter. This parameter describes a subject's response strategy. While *gamma* > 1 allows for similarity models to account for deterministic responses, *gamma* < 1 allows to account for probabilistic responses. Because our procedure involves probabilistic feedback to fit the data, both GCM and MPM converge on *gamma* values lower than 1. As Nosofsky and Zaki (2002) discuss, the *gamma* parameter is added to the choice response equation (see eq. (3)), where each similarity (i.e., $S_A(x)$ and $S_B(x)$) includes the *gamma* scaling-parameter. However, because our ALF model does not predict the p(category | exemplar) by using Luce's axiom, we added a *g* parameter to the logistic function (see eq. (6)). The *g* parameter in the ALF has the same assumptions and constraints as the *gamma* parameter. Indeed, partial correlations controlling by condition confirm this assumption. The *g* parameter estimated through modeling is significantly correlated to MPM *gamma* ($r = .57$, $p < .001$) and GCM *gamma* ($r = .98$, $p < .001$). This shows that our *g* parameter behaves similarly to the *gamma* parameter of similarity-based models.

Discussion

In the current work, we contrasted two well-known similarity-based models (GCM and MPM) with a model based on associative learning strength (ALF) under conditions of uncertainty (i.e., probabilistic classification). Our review of the literature reveals that procedures that explore category learning under uncertainty are sparse (Marchant & Chaigneau, 2021). Because natural learning scenarios and categorization of real-life objects and events rarely occur in an invariant context (cf., Tversky & Kahneman, 1974; Estes, 1976), it is interesting to test traditional similarity-based models' ability to account for data under these conditions. To this end, we trained subjects in a probabilistic category learning task that uses different exemplar probabilities by adjusting the p(category | exemplar).

Fitting computational models to empirical data, enables direct comparison between competing models (Wills & Pothos, 2012). When performing those comparisons, our results show that under a probabilistic category learning paradigm our ALF model generally produces better fits of subjects' category learning data. The reasons why similarity-based models perform worse in a probabilistic category learning environment are complex. A first element to consider is that successful performance in probabilistic categorization is far from optimal performance in deterministic categorization. As shown in Figure 3, the exemplar that produced the best performance in our experiment, produced only 77 % of correct classifications (Exemplar E1 in condition 2). Because in experiments that use a DF response criterion, feedback is consistent throughout the complete experiment, people can achieve perfect performance. In contrast, in PF experiments perfect performance is not possible. However, people should be near the "optimal" probabilistic performance set by the experimenters (Shanks et al., 2002). To understand why the GCM is not equipped to handle probabilistic performance, it is useful to examine Eq. (1) in more detail. That equation reflects that the GCM assumes that each exemplar is compared to all the other available ones. Thus, the GCM extracts a virtual prototype (i.e., the category centroid) from the data. Because, though learning had already occurred in block 3 (see Figure 2), subjects' performance was still highly variable during that block, the GCM could not extract category structure from that data, leading to the undifferentiated w parameters in Table 5 and Figure 4. For preliminary evidence consistent with our explanation, see Rouder and Ratcliff (2004).

The situation is different for the MPM for precisely the same reasons discussed above. In contrast to the GCM, the MPM requires being given the category prototype and does not extract it solely from the data (as a reminder, prototypical category A was set to "111"). Inspecting Eq. (4) helps making this point. The prototype that figures in Eq. (4) does not figure in Eq. (1). This allows the MPM to extract information from the data even under noisy conditions such as those in our experiment.

However, the ALF model outperforms both similarity-based models. The reasons for it being better than the GCM were discussed above. It is striking though that the AFL outperforms the MPM. Similar to the prototype model, which is provided with the prototype, the experimenter provides the AFL with the feature weights implied in the probabilistic task structure, which suggests that they might perform similarly. Furthermore, Nosofsky (1992) suggested that the AFL is equivalent to a prototype model, which also hints that they should perform similarly. However, as we have shown, this is not the case. Furthermore, the AFL achieves its fit with a single free parameter (g) while the MPM requires 4 free parameters and does not achieve the same fits attained by the AFL. We believe the main reason why the AFL outperforms the other models must be because of the correlation coefficients in Table 3, which are fed into Eq. (6) for our modeling, closely reflect the task structure that subjects experienced in our task. Recall that the weights shown in Table 3 are the correlation coefficients relating each individual feature with the category under the uncertain feedback conditions. This is precisely what the learning algorithm in Eq. (5) holds, i.e., that people learn those associations through feedback. This strongly suggests that under conditions of probabilistic feedback, people rely on an associative learning mechanism (perhaps motor-driven), rather than on the elaboration of a logical verbal rule in which declarative memory is involved.

Prima facie, our modeling effort shows that similarity computations may not be the best way to conceptualize category learning in PF tasks. If not similarity, which computations might be apt? Continuing with the topic we briefly discuss in the introductory section, we propose that category learning with PF is a procedural learning process, and that ALF models it. Consider that ALF is based on previous connectionist learning models (component-cue network model; Gluck & Bower, 1988) which use an error metric that converges to a least mean squared (LMS) solution through the association of the input patterns with their outputs. Note, that Eq. (5) is the delta rule first proposed by Widrow and Hoff (1960), which both ALF and connectionist models used to account for a LMS solution. It is generally accepted that this delta rule is similar

to the Rescorla and Wagner (1972) model of animal associative learning. However, the proposal in the current work differs from a connectionist learning model on how the model estimates p(category | exemplar) (i.e., the logistic function in Eq. (6)).

Continuing with the procedural learning theme, note that there is a large body of research linking procedural learning with the Basal Ganglia (Ashby & Spiering, 2004; Ashby & Ennis, 2006; Ashby, Ennis, & Spiering, 2007). Thus, and consistently with findings in clinical patients using the Weather Prediction Task (Knowlton et al., 1994; Gluck et al., 1996), we want to hypothesize that the brain mechanisms that are being modeled by the ALF may be implemented in cortico-cortical loops involving the Basal Ganglia (Lawrance et al., 1998; Seger, 2006; 2008). Future studies could further explore ALF model predictions regarding Basal Ganglia activation by employing neuroimaging techniques or neuropsychological participants.

A final issue worth considering is model parsimony. In general, given similar explanatory power, the simpler model is to be preferred. On that ground, ALF is also superior to the similarity-based counterparts. Our results show that the ALF achieves better fits (i.e., more explanatory power), even before penalizing models for their number of free parameters (i.e., parsimony).

Future directions

Based on results like those reported here, work in our laboratory has recently focused on the ALF as a model of procedural category learning model. Several issues remain to be explored. We do not discuss them in depth here, but only provide a list so the reader can sense the research program's potential. An initial question we would like to address is whether we could combine a procedural-based model with a similarity-based model. A previous well-known model has endeavored to accomplish this purpose. The Attentional Learning COVEring map (ALCOVE; Krushcke, 1992) is a connectionist model which combines exemplar category representation with an error-driven metric (i.e., the delta rule discussed above). A logical follow-up is to compare ALCOVE and ALF under conditions of uncertainty. Note that in the current work we have shown that a purely procedural-based model could account for probabilistic category learning without implying a similarity metric. However, further evidence and comparisons with other models such as ALCOVE would be desirable.

A related issue in computational models of categorization is the relevance for models to capture individual learning trajectories (Ashby; Maddox, & Lee, 1994; Shen & Palmeri, 2016). Accounting for individual learning curves would reveal different time-points where each participant achieves an optimal performance. Preliminary work in our laboratory shows that the ALF can capture individual trajectories during learning and predict individual classification predictions. Other phenomena that the ALF should be able to handle, which open alternative for future work, are the base-rate neglect phenomena (Estes et al., 1989), feedback discounting (Craig et al., 2011) and feature-blocking (Bott et al., 2007).

In sum, here we provided evidence for a simple procedural-based model of category learning that efficiently accounts for people's learning performance under conditions of uncertainty. We expect these contributions to be meaningful for computational psychology in general and for cognitive modelers who explore the quantitative implementation of categorization theories. Furthermore, the mathematical formulations of the ALF may also be interesting to other fields in the Cognitive Sciences, such as AI.

References

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., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation - Advances in Research and Theory, 46*, 1–36. https://doi.org/10.1016/S0079-7421(06)46001-1

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*(3), 632–656. https://doi.org/10.1037/0033-295X.114.3.632

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

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology,* 56(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*(1), 147–161. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, *5*(3), 144–151. https://doi.org/10.1111/j.1467-9280.1994.tb00651.x

Ashby, F. G., & Spiering, B. J. (2004). The neurobiology of category learning. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 101–113. https://doi.org/10.1177/1534582304270782

Ashby, G. F., & Valentin, V. V. (2017). Multiple systems of perceptual category learning: Theory and cognitive tests. In Cohen, H., & Lefebvre, C. (Eds). *Handbook of Categorization in Cognitive Science*. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-101107-2.00007-5

Ashby, F. G., & Valentin, V. V. (2018). The categorization experiment: Experimental design and data analysis. In Wagenmakers, E. J., & Wixted, J. T. (Eds). *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–41). John Wiley & Sons. https://doi.org/10.1002/9781119170174.epcn508

Bott, L., Hoffman, A. B., & Murphy, G. L. (2007). Blocking in category learning. *Journal of Experimental Psychology: General*, *136*(4), 685–699. https://doi.org/10.1037/0096-3445.136.4.685

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory and Cognition*, *30*(3), 353–362. https://doi.org/10.3758/BF03194936

Craig, S., Lewandowsky, S., & Little, D. R. (2011). Error discounting in probabilistic category learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 673–687. https://doi.org/10.1037/a0022473

D'Errico, J. (2021). fminsearchbnd, fminsearchcon, MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon

Ell, S. W., Smith, D. B., Peralta, G., & Hélie, S. (2017). The impact of category structure and training methodology on learning and generalizing within-category representations. *Attention, Perception, and Psychophysics*, *79*(6), 1777–1794. https://doi.org/10.3758/s13414-017-1345-2

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, *83*(1), 37–64. https://doi.org/10.1037/0033-295X.83.1.37

Estes, W. K., Campbell, J. A., Hatsopoulos, N., & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 556–571. https://doi.org/10.1037/0278-7393.15.4.556

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*(3), 227–247. https://doi.org/https://doi.org/10.1037/0096-3445.117.3.227

Gluck, M. A., Oliver, L. M., & Myers, C. E. (1996). Late-training amnesic deficits in probabilistic category learning: A neurocomputational analysis. *Learning and Memory*, *3*(4), 326–340. https://doi.org/10.1101/lm.3.4.326

Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General*, *139*(2), 319–340. https://doi.org/10.1037/a0019042

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399–1402. https://doi.org/10.1126/science.273.5280.1399

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning Memory*, *1*(2), 106–120. https://doi.org/10.1101/lm.1.2.106

Knowlton, B. J., Swerdlow, N. R., Swenson, M., Squire, L. R., Paulsen, J. S., & Butters, N. (1996). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology, 10*(4), 538–548. https://doi.org/10.1037/0894-4105.10.4.538

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.

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General*, *135*(2), 162–183. https://doi.org/10.1037/0096-3445.135.2.162

Lawrence, A. D., Sahakian, B. J., & Robbins, T. W. (1998). Cognitive functions and corticostriatal circuits: insights from Huntington's disease. *Trends in Cognitive Science*, *2*(10), 379–388. https://doi.org/10.1016/S1364-6613(98)01231-5

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Sage.

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

Marchant, N., & Chaigneau, S. E. (2021). Designing probabilistic category learning experiments: The probabilistic prototype distortion task [Paper presentation]. 43rd Annual Meeting of the Cognitive Science Society: Comparative Cognition: Animal Minds, CogSci 2021, Virtual, Austria. https://pure.uai.cl/en/publications/designing-probabilistic-category-learning-experiments-the-probabi

Meeter, M., Radics, G., Myers, C. E., Gluck, M. A., & Hopkins, R. O. (2008). Probabilistic categorization: How do normal participants and amnesic patients do it? *Neuroscience and Biobehavioral Reviews, 32*(2), 237–248. https://doi.org/10.1016/j.neubiorev.2007.11.001

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

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, 1*, 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*(1), 25–53. https://doi.org/10.1146/annurev.ps.43.020192.000325

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In Photos, E. M., & Wills, A. J. (Eds). *Formal Approaches in Categorization* (pp. 18–39). Cambridge. https://doi.org/10.1017/cbo9780511921322.002

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., & 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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rehder, B., Colner, R. M., & Hoffman, A. B. (2009). Feature inference learning and eyetracking. *Journal of Memory and Language*, *60*(3), 393–419. https://doi.org/10.1016/j.jml.2008.12.001

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64–99). Appleton-Century-Crofts.

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

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, *133*(1), 63–82. https://doi.org/10.1037/0096-3445.133.1.63

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Seger, C. A. (2006). The basal ganglia in human learning. *Neuroscientist, 12*(4), 285–290. https://doi.org/10.1177/1073858405285632

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*, *32*(2), 265–278. https://doi.org/10.1016/j.neubiorev.2007.07.010

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*(1), 203–219. https://doi.org/10.1146/annurev.neuro.051508.135546

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250. https://doi.org/10.1002/bdm.413

Shen, J., & Palmeri, T. J. (2016). Modelling individual difference in visual categorization. *Visual Cognition*, *24*(3), 260–283. https://doi.org/10.1080/13506285.2016.1236053

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

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience & Biobehavioral Reviews*, *32*(2), 219–236. https://doi.org/10.1016/j.neubiorev.2007.07.008

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, 4*, 96–194.

Widrow, B., & Kamenetsky, M. (2003). Statistical efficiency of adaptive algorithms. *Neural Networks*, *16*(5–6), 735–744. https://doi.org/10.1016/S0893-6080(03)00126-6

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

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning Memory and Cognition*, *28*(3), 585–593. https://doi.org/10.1037/0278-7393.28.3.585

Fecha de recepción: Junio de 2021.
Fecha de aceptación: Abril de 2022.