1 Generating Exemplars from Categories-by-Negation

When generating exemplars from categories defined by a label that describes what is *not* already learned, we can imagine a mechanism that focuses on the areas of the feature space that is not occupied by learned categories. This can be implemented with a simple model where exemplars from these novel categories are generated with a probability inversely proportional to the density of all other categories. For simplicity, we assume that each individual category has its exemplars distributed as a multivariate Gaussian parameterized by a d-sized vector of empirical means M and a d-by-d empirical covariance matrix S, where d describes the dimensionality (i.e., number of features).

More formally, the probability of generating an exemplar y given a set of categories C'[1] excluding the novel category can be described as proportional to a density function $n(y|C')$ where

$$n(y|C') = -\sum_{c \in C'} \mathrm{MvN}(y; M_c, S_c) \tag{1}$$

To get the actual choice probabilities, we normalize these values using an exponentiated Luce's choice rule:

$$p(y|C') = \frac{\exp(\theta \cdot n(y|C'))}{\sum_i \exp(\theta \cdot n(y|C'))} \tag{2}$$

where θ is a response determinism parameter that is freely estimated.

Note that with this simple model the probability of generating an exemplar is completely dependent on the learned categories – there is no explicit within-category constraints defined by the model.

Although this model is limited in its complexity, it is straightforward to apply as an extension to any of the other models of category generation. We describe extensions to four models – PACKER, copy-and-tweak, Jern and Kemp's (2013) hierarchical Bayesian model, and the representativeness model – in the following sections.

2 Extending PACKER and copy-and-tweak

We begin by briefly describing the structure of the PACKER and copy-and-tweak models (which can be implemented as a special case of PACKER).

These models predict the generation of novel exemplars by measuring the candidate exemplar's similarity to other exemplars. Specifically, the similarity between exemplars x_i and x_j is computed using Shepard's law:

[1] This model assumes that the novel category is a negated category with respect to *all* learned categories. It would be appropriate for data from people generating exemplars that are not what they previously learned. However, if participants were instead instructed to generate exemplars that are not from some subset of a larger set of learned categories (e.g., generate not Alphas, having learned both Alpha and Beta category exemplars), this simple model would not be appropriate. For our experiments so far, we only test participants who have learned one category, so there is no problem at this point. For future development, however, we might want to consider negative-space models that can consider a subset of learned categories. This is probably as simple as having C' represent the relevant subset of learned categories.

$$s\left(x_i, x_j\right) = \exp\left\{-c\left[\sum_k w_k \left|x_{ik} - x_{jk}\right|^r\right]^{1/r}\right\} \tag{3}$$

where k indicates each feature, w_k is the weight placed on that feature, c is a free-to-vary specificity parameer, and r allows us to define the distance metric. In most of our simulations, we fix r to 1 (implementing a city-block distance metric). Unless otherwise specified, w_k is fixed to be equal across all features.

PACKER takes into account the similarity of a candidate exemplar to the existing novel category (i.e., the target category) exemplars as well as its dissimilarity to contrast category exemplars. This is implemented as an aggregated similarity a between candidate y and all other exemplars x:

$$a(y, x) = \sum_j f(x_j) s(y, x_j) \tag{4}$$

where $f(x_j) = \theta_t$ when x_j is a member of the target category and $f(x_j) = \theta_c$ when x_j is a member of a contrast category.

To extend this model with the negative-space mechanism, we linearly combine this aggregated similarity with the densities of the contrast categories $n(.|C')$ (Equation 1):

$$b(y, x) = a(y, x) + \gamma \cdot n(y|C') \tag{5}$$

where γ is a free-to-vary parameter that describes the amount of weight placed on the negative-space mechanism.

The probability that a given item y will be generated given the model's memory x is computed using relative summed similarity values across all generation candidates y_i (i.e., with an exponentiated Luce's choice rule):

$$p(y) = \frac{\exp\left\{b\left(y, x\right)\right\}}{\sum_i \exp\left\{b\left(y_i, x\right)\right\}} \tag{6}$$

The negative-space copy-and-tweak model is instantiated in exactly the same way, but with θ_c set to 0, since the model does not account for contrast category similarity. In addition, θ_t and γ combine to serve as the response determinism parameter that is free to vary. Response determinism for PACKER is a combination of θ_c, θ_t, and γ parameters.

3 Extending Hierarchical Bayesian and Representativeness models

Unlike the exemplar models, the hierarchical Bayesian and Representativeness models do not explicitly measure exemplar similarity on the feature space. Instead, both of these Bayesian models assume exemplars x within a category c are influenced by an underlying multivariate Gaussian distribution parameterized by means μ_c and covariance matrix Σ_c.

μ_c and Σ_c are assumed to be Normal-Inverse-Wishart distributed. The former can be straightforwardly computed with:

$$\mu_c = \frac{\kappa\mu_0 + n_c\bar{x}_c}{\kappa + n_c} \tag{7}$$

where:

- μ_0 is the prior mean along d dimensions. Here we set it to the middle of the space.
- κ is a scalar hyper-parameter, roughly weighting the importance of μ_0. κ must be greater than zero.
- n_c is the number of observations in x_c
- $\bar{x}_c$ is the sample mean along d dimensions

Σ_c is influenced by members of the target category c as well as all other members through a domain-level covariance matrix Σ_D. Assuming the observed category covariances S_c are Wishart-distributed, we can compute Σ_D with:

$$\Sigma_D = \Sigma_0 + \sum_c S_c \tag{8}$$

where Σ_0 is a d-by-d prior covariance matrix defined as the d-dimensional identity matrix I_d weighted by a free parameter *lambda*:

$$\Sigma_0 = \lambda I_d \tag{9}$$

With this, we can compute Σ_c as:

$$\Sigma_c = [\Sigma_D\nu + S_c + \frac{\kappa n_c}{\kappa + n_c}(\bar{x_c} - \mu_c)(\bar{x_c} - \mu_c)^T](\nu + n_c)^{-1} \tag{10}$$

where ν is an additional free parameter weighting the importance of Σ_D.

As with the other models, choice probabilities are computed using the exponentiated Luce's choice rule. With some general function $g(\cdot)$:

$$p(y) = \frac{\exp\{\theta \cdot g(\cdot)\}}{\sum_i \exp\{\theta \cdot g(\cdot)\}} \tag{11}$$

The specific form of $g(\cdot)$ is where the hierarchical Bayesian model diverges from the representativeness model. For the hierarchical Bayesian model , $g(\cdot)$ is the density at y under a multivariate Gaussian parameterized by μ_c and Σ_c:

$$g(\cdot) = \text{Normal}(y; \mu_c, \Sigma_c) \tag{12}$$

For the representativeness model, $g(\cdot)$ is the representativeness $R(y, h)$ of y from the target hypothesis h compared to all other hypotheses h'. Here, each hypothesis is represented by the probability density of the underlying distribution of a specific category. Formally,

$$\begin{align} g(\cdot) &= R(y, h) \tag{13} \\ &= \log \frac{p(y|h)}{\sum_{h' \in \mathcal{H}^c} p(y|h')p(h')}. \tag{14} \end{align}$$

where h_c is a multivariate Gaussian parameterized by μ_c and Σ_c for a specific category c and $\mathcal{H}^{\rfloor}$ represents the hypothesis set from all contrast categories.

Similar to the exemplar models, the extension of both hierarchical Bayesian and the representativeness models are performed by linearly combining the negative-space densities computed from $n(y|C')$ to the original function prior to normalization via the exponentiated Luce choice rule. Formally, we can represent the negative-space extensions of both Bayesian models by amending Equation 11 to:

$$p(y) = \frac{\exp\{\theta \cdot g(\cdot) + \gamma \cdot n(y|C')\}}{\sum_i \exp\{\theta \cdot g(\cdot) + \gamma \cdot n(y|C')\}} \tag{15}$$

Given the structure of the Bayesian models, we adopt a more principled implementation of $n(y|C')$ by specifying μ_c (from Equation 7) and Σ_c (from Equation 10) as the means and covariance matrices of the multivariate Gaussian, instead of the empirical M_c and S_c values. Specifcally, Equation 1 becomes:

$$n(y|C') = -\sum_{c \in C'} \mathrm{MvN}(y; \mu_c, \Sigma_c) \tag{16}$$