PACKER: An Exemplar Model of Category Generation

Nolan Conaway (nconaway@wisc.edu)
Joseph L. Austerweil (austerweil@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

The creative generation of new concepts is an intriguing yet understudied topic in cognitive science. In this paper, we present a novel exemplar-based approach to category generation: PACKER (*Producing Alike and Contrasting Knowledge using Exemplar Representations*). PACKER's core design assumptions are that (*a*) categories are represented as exemplars in a multidimensional psychological space, (*b*) generated items should be similar to exemplars of the same category, and (*c*) generated categories should be dissimilar to existing categories. A behavioral study and formal simulations reveal broad support for PACKER's predictions.

Keywords: Categorization, exemplar models, category generation, creative cognition.

Introduction

The creation of new concepts and ideas is among the most interesting – yet infrequently studied – capabilities of human cognition. This paper focuses on one topic within the broader field of creative cognition: category generation. Foundational work on this topic (e.g., Smith, Ward, & Schumacher, 1993; Ward, 1994, 1995; Ward, Patterson, Sifonis, Dodds, & Saunders, 2002) has focused on the role of prior knowledge in generating novel concepts. A core phenomenon is that people generate categories with similar distributional properties as existing categories. For example, Ward (1994) asked participants to generate species of plants and animals that might exist on other planets. Generation was strongly constrained by prior knowledge of Earth species: People generated species with the same features as those found on Earth (e.g., eyes, legs, wings) and possessing the same feature correlations observed on Earth (e.g., feathers co-occur with wings).

Recent work has proposed and tested computational models of the processes underlying category generation. Jern and Kemp (2013) trained participants on a set of experimenter-defined categories composed of exemplars within an artificial three-dimensional domain (e.g., 2D shapes varying in size, hue, and saturation). After a short training phase, participants were asked to generate exemplars from a new category. Participants were provided with a set of scales to adjust the feature values of each generated stimulus, and were given unlimited time to create each example. As in the classic Ward (1994) experiment, Jern and Kemp (2013) found that generated categories possessed the same feature variance and correlations as the experimenter-defined categories in the domain. They evaluated different possible computational models of how people generate categories (such as an exemplar-based "copy-and-tweak" model) and found that a hierarchical Bayesian model (which generates categories by abstracting the distributional structure of the domain) provided the strongest account of the behavioral results.

In this paper, we introduce a novel exemplar-based approach to category generation, PACKER (*Producing Alike and Contrasting Knowledge using Exemplar Representations*), which creates categories by balancing two constraints: (1) new categories should be different from known categories (minimizing between-class similarity), and (2) new categories should be internally coherent (maximizing within-class similarity). As such, PACKER is a significant departure from previous accounts of generation – rather than proposing that people create categories by abstracting and re-using knowledge of related categories, PACKER first considers how the generated category should *differ* from related categories. Further, it does so using the well-studied mechanics of exemplar representations and therefore possesses a rich connection to the wider body of research on category learning.

In the sections below, we formally describe the PACKER model and explore its predictions in a behavioral experiment. We compare its performance to a copy-and-tweak and hierarchical Bayesian model by examining their fits to aggregate results and individual differences.

PACKER: An Exemplar Model

The PACKER model is an extension of the prominent Generalized Context Model of category learning (GCM; Nosofsky, 1984). The model assumes that each category is represented by a collection of exemplars within in a k-dimensional psychological space, and that generation is constrained by both similarity to members of the target category (the category in which a stimulus is being generated) as well as similarity to members of other categories.

As in the GCM, the similarity between two examples, $s(x_i, x_j)$, is an inverse-exponential function of distance:

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

where w_k is the attention weighting of dimension k ($w_k \geq 0$ and $\sum_k w_k = 1$), accounting for the relative importance of each dimension when calculating similarity, and c ($c > 0$) is a specificity parameter controlling the spread of exemplar generalization. For simplicity, most simulations will use uniform attention weights. We will use non-uniform attention weights in our discussion of individual differences.

To generate a new example, the model considers both the similarity to examples from contrast categories as well as similarity to examples in the target category. The aggregated similarity a between generation candidate y and the model's stored exemplars x can be computed as $a(y, x) =$

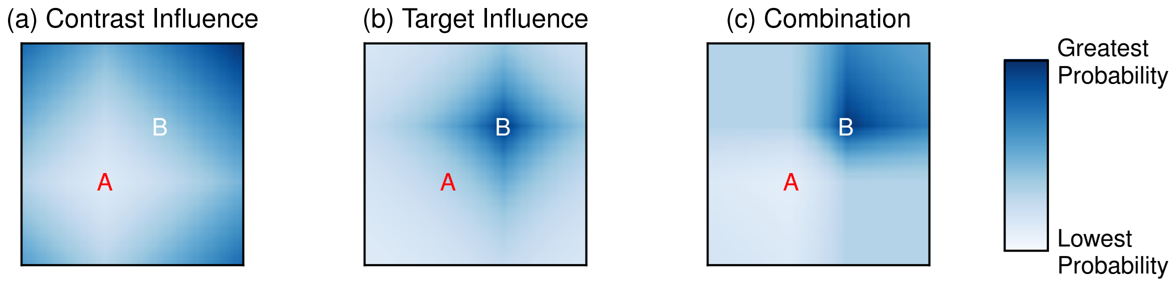

Figure 1: PACKER generation of a category 'B' example, following exposure to one member of category 'A' and one member of category 'B'. *(a)* Predictions based on contrast similarity only $\{\phi = -1, \gamma = 0\}$. *(b)* Predictions based on target similarity only $\{\phi = 0, \gamma = 1\}$. *(c)* Predictions with both constraints considered $\{\phi = -1, \gamma = 1\}$.

$\sum_j f(x_j)s(y,x_j)$, where $f(x_j)$ is a function specifying the extent to which each exemplar contributes to the target category. For brevity, we will abbreviate $f(x_j)$ as f_j. PACKER sets f_j depending on exemplar j's category membership: $f_j = \phi$ if x_j is a member of the contrast category, and $f_j = \gamma$ if x_j is a member of the target category. ϕ and γ are free parameters ($-\infty \leq \phi, \gamma \leq \infty$) controlling the contribution of contrast- and target-category similarity, respectively. Larger absolute values for either parameter produce greater consideration of those exemplars, with values of 0 producing no effect. A negative value for f_j produces a 'repelling' effect (exemplars are less likely to be generated nearby x_j). Conversely, a positive value for f_j produces a 'pulling' effect (exemplars are more likely to be generated nearby x_j).

Recall that we proposed that categories are generated by balancing two constraints: that new categories should be different from existing categories, and exemplars belonging to the same category should be similar to one another. This is realized in PACKER when $\phi < 0$, and $\gamma > 0$. Negative values for ϕ encourage y to be distant from the contrast category (as similarity to contrast category exemplars are subtracted during aggregation). Positive values for γ encourage y to be close to other exemplars of the target category. When $|\phi| = \gamma$, the "force" of repulsion from the contrast category is equal to the "force" of attraction to the target category.

The probability that a given candidate y will be generated is evaluated using an Exponentiated Luce (1977) choice rule. Candidates with greater values of a are more likely to be generated than candidates with smaller values:

$$p(y) = \frac{\exp(\theta \cdot a(y,x))}{\sum_i \exp(\theta \cdot a(y_i,x))} \tag{2}$$

where θ ($\theta \geq 0$) controls response determinism.

Summary

The proposed PACKER model suggests people generate categories by minimizing between-category similarity and maximizing within-category similarity. The underlying processes assumed by PACKER are highly similar to those in the GCM, with the only alteration being that PACKER aggregates positive- and negative-valued similarities, rather than only aggregating positive-valued similarities.

In later sections, we will explore the unique predictions yielded by these design principles. First, however, we contrast PACKER with other category generation models.

Previous Accounts of Category Generation

Previous models of category generation focus on capturing the tendency for people to produce new categories that have similar distributional properties to existing categories. To the best of our knowledge, Jern and Kemp (2013) were the first to quantitatively evaluate computational models of generation. Based on their work, we describe two alternative models: a formalization of the *copy-and-tweak* hypothesis (Ward, 1995; Ward et al., 2002), and the *hierarchical sampling* proposed by Jern and Kemp (2013).

Copy-and-Tweak

The *copy-and-tweak* model proposes that generation is a two-part process: First, learners retrieve an observation of the target class from memory, and then they tweak it make something new. Jern and Kemp (2013) interpreted this proposal in terms of an exemplar model using the GCM (Nosofsky, 1984). Formally, their model is equivalent to PACKER with $\{\phi = 0, \gamma = 1\}$ (see Figure 1). To allow for more flexible performance, and to explicitly formalize the 'tweak' procedure, we slightly generalized their implementation.

In our copy-and-tweak model, all observations z from the target class are equally likely to be retrieved as the source: $p(z) = 1/n$, where n is the number of stored examples in the target class. If $n = 0$, learners are assumed to generate an item at random. After a source exemplar is retrieved, the similarity between generation candidates y and the retrieved exemplar is computed as per Equation 1. The model's goal is to generate an item that is similar to z, but does not exceed a given similarity threshold: The generated item should be similar, but not too similar to z. Formally, the probability that candidate exemplar y will be generated based on a source exemplar z is

$$p(y|z) = \frac{\exp\{\theta \cdot s(y,z)\} I\left(s(y,z) \leq \tau\right)}{\sum_i \exp\{\theta \cdot s(y_i,z)\} I\left(s(y_i,z) \leq \tau\right)} \tag{3}$$

where θ is a response determinism parameter, τ is the similarity threshold, and $I(\cdot)$ is the indicator function, which returns 1 when it is passed a true expression and 0 otherwise. So, $I(s(y,z) \leq \tau)$ is 1 when the candidate exemplar y is far enough away from the source exemplar z. This ensures the source exemplar is tweaked "enough". When $\tau = 1$, the threshold has no effect, and our model is equivalent to that of Jern and Kemp (2013). To obtain predictions not depending on a given source example, the model's predictions can be aggregated over all possible sources: $p(y) = \sum_z p(z)p(y|z)$.

Hierarchical Sampling

Based on several results inconsistent with the copy-and-tweak account, Jern and Kemp (2013) advocated a hierarchical Bayesian model. Exemplars of each category were generated from a multivariate Gaussian distribution over the dimensions of stimulus space. The mean of each category was independent, but the covariance matrices (encoding feature variances and correlations) were generated from a common prior distribution. New categories are generated then by generating a new mean (uniform over stimulus space) and covariance matrix from the common prior distribution. Because the shared prior distribution's parameters were unobserved, a hierarchical Bayesian model uses information from the previous categories (their feature variances and correlations) to generate the covariance matrix of the new category.

We assume that each category's exemplars are distributed according to a multivariate normal distribution with parameters (μ, Σ). We assume that each category's covariance matrix is inverse-Wishart distributed with parameters (ν, κ, and Σ_D).[1] Σ_D is the covariance matrix shared between categories. We assumed the shared covariance matrix Σ_D was generated from a Wishart distribution (for conjugacy) with parameters ν_0, κ_0, and Σ_0. We set $\nu_0 = \kappa_0 = 1$, and $\Sigma_0 = \rho\mathbf{I}$, where ρ was a free parameter controlling the expected variance of dimensions (dimensions of the shared covariance matrix are expected to be uncorrelated) and $\mathbf{I}$ is the identity matrix.

To simplify the model predictions, we used *maximum a posteriori* (MAP) estimates for the hidden parameters and then generated new categories based on those estimates. Due to conjugacy, the MAP estimate for the shared covariance matrix $\Sigma_D = \Sigma_0 + \sum_c C_c$, where C_c is the empirical covariance matrix of category c. The MAP estimate of the covariance matrix for category B is

$$\Sigma_B = \left[\Sigma_D \nu + C_B + \frac{\kappa n_B}{\kappa + n_B}(\bar{x_B} - \mu_B)(\bar{x_B} - \mu_B)^T\right](\nu + n_B)^{-1} \quad (4)$$

where ν ($\nu > k-1$) is an additional free parameter (from the Inverse-Wishart prior on Σ_B) weighting the importance of Σ_D. When the target category has no members (i.e., $n_B = 0$), items are generated at random.

[1] Note that Jern and Kemp (2013)'s model is slightly different, as they used a non-conjugate model. Their model acts very similar to our version of it and receives comparable fits.

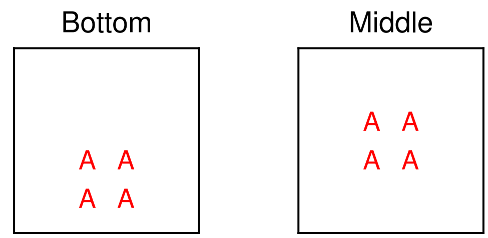

Figure 2: Conditions tested in the behavioral experiment.

Generated exemplars are then assumed to be drawn from the multivariate normal distribution specified by (μ_B, Σ_B). Thus, $p(y)$ is

$$p(y) = \frac{\exp(\theta \cdot \text{Normal}(y; \mu_B, \Sigma_B))}{\sum_i \exp(\theta \cdot \text{Normal}(y_i; \mu_B, \Sigma_B))} \quad (5)$$

where θ is a response determinism parameter and $\text{Normal}(y; \mu, \Sigma)$ denotes a multivariate Normal density evaluated at y.

Behavioral Experiment

The copy-and-tweak and hierarchical sampling models were designed to explain effects of prior knowledge on the structure of categories, and so these models assume little about where in stimulus space categories are likely to be generated. Indeed, when there are no known examples of the target category, both models assume that generation is random. PACKER is thus unique in its prediction that contrast categories should influence both the structure and location of generated categories. The behavioral experiment described below was designed to test this key prediction.

The experiment follows the paradigm developed by Jern and Kemp (2013): first, participants learn members of a known category ('Alpha', or 'A'), and are then asked to generate exemplars belonging to a new category ('Beta', or 'B'). We developed two Alpha categories (see Figure 2): the 'Bottom' Alpha category is a tight cluster in the bottom-center of the space, and the 'Middle' Alpha category is identical except that it lies in the center of stimulus space. The two categories only differ in their location, not in distributional structure.

Although our manipulation is minimal, the PACKER model is capable of predicting strong between-condition differences. PACKER proposes that the nature of the space not occupied by the Alpha category determines where members of the Beta category are likely to be generated. Thus, the lower areas of the stimulus space should be less frequently used for generation in the Bottom condition compared to the Middle (as these areas possess greater similarity to the Bottom Alpha category). Conversely, the upper areas of the stimulus space should be used for generation more frequently in the Bottom condition compared to Middle.

More generally, PACKER proposes that the probability a stimulus y will be generated is a function of its similarity to contrast categories *and* to members of the target category.

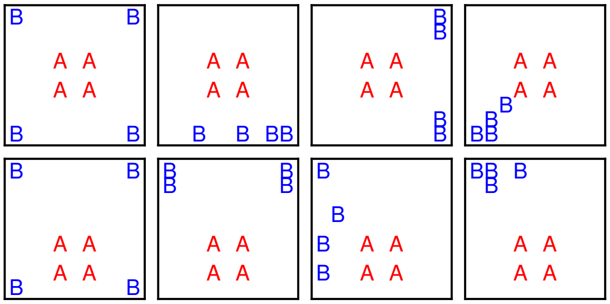

Figure 3: Sample generated categories.

Two more general predictions (not specific to either condition) follow from this proposal: (1) the location of Beta examples should be positively related to distance from the Alpha category, and (2) Beta examples should be more similar to one another than they are to members of the Alpha category.

Participants & Materials We recruited 122 participants from Amazon Mechanical Turk from the US equally assigned to each condition. Stimuli were squares varying in color (RGB 25–230) and size (3.0–5.8cm). The assignment of perceptual features (color, size) to axes of the domain space (x, y) was counterbalanced across participants.

Procedure Participants began the experiment with a short training phase (3 blocks of 4 trials), where they observed exemplars belonging to the 'Alpha' category. Participants were instructed to learn as much as they can about the Alpha category, and that they would answer a series of test questions afterwards. On each trial, a single Alpha category exemplar was presented, and participants were given as much time as they desired before moving on. Exemplars were randomly ordered within each block. Participants were shown the range of possible colors and sizes prior to training.

Following the training phase, participants were asked to generate four examples belonging to another category called 'Beta'. Participants were instructed that members of the Beta category could be quite similar or different depending on what they think makes the most sense for the category, but that they were not allowed to make the same example twice. As in Jern and Kemp (2013), generation was completed using a sliding-scale interface. Participants were presented with two scales that controlled the dimensions of the generated example. An on-screen preview of the example updated whenever one of the features was changed. Participants could generate any example along an evenly-spaced 9x9 grid, except for any previously generated Beta exemplars. Neither the members of the Alpha category nor the previously generated Beta examples were visible during generation.

Results Because the conditions differ only in their location along the y-axis (see Figure 2), we focus on how Beta exemplars are generated above and below the contrast category. Several sample Beta categories are depicted in Figure 3. As is

Table 1: Behavioral results.

Middle	Used top row	No top row
Used bottom rows	28	18
No bottom row	11	4
Bottom	Used top row	No top row
Used bottom row	16	8
No bottom row	31	6

evident in Figure 3, we observed broad individual differences in generation strategy: Whereas some participants generated all four Beta examples within a narrow y-axis range, others generated Beta examples along a wide range.

To evaluate the key predictions of PACKER, we determined the number of participants in each condition who placed at least one Beta examplar on the top and bottom 'rows' of the space (the maximum and minimum possible y-axis value, respectively). The resulting contingencies data are shown in Table 1. Fisher's Exact Tests reveal that more Middle participants generated a Beta exemplar in the bottom row , $p < 0.001$, but the conditions did not differ in use of the top of the space, $p = 0.16$. More Middle participants placed Beta exemplars in the top *and* bottom rows, $p = 0.038$.

To evaluate PACKER's other predictions, we computed the number of exemplars produced at different distances to the center of the Alpha category. These data, depicted in the left panel of Figure 4 reveal a strong preference for stimuli that are dissimilar to the members of the Alpha category: maximally distant items were by far the most frequently generated.

Finally, we computed for each participant the average distance between exemplars belonging to the same and opposite categories. These data (depicted in the right panel of Figure 4) show that, as observed by Ward (1994), most people generated Beta categories in which members are closer to one another than they are to members of the Alpha category (i.e., more between- than within-category distance). We did however, observe a notable subset of individuals with greater within-class distance. These individuals tended to adopt a 'corners' approach, in which a Beta examples were placed almost exclusively in the corners of the space.

Summary

Our results support PACKER's predictions: People tend to generate items that are dissimilar from the contrast category and similar to the target category (with some exceptions, see Figure 4). By consequence, we observed considerable differences in generation between the Middle and Bottom conditions: participants in the Bottom condition were less likely to use the bottom row of the stimulus space for generation, and participants in the Middle condition were more likely to create categories spanning the entire y-axis (utilizing the top and bottom row of the space). This latter result is especially interesting as it conflicts with previous results: Qualitatively

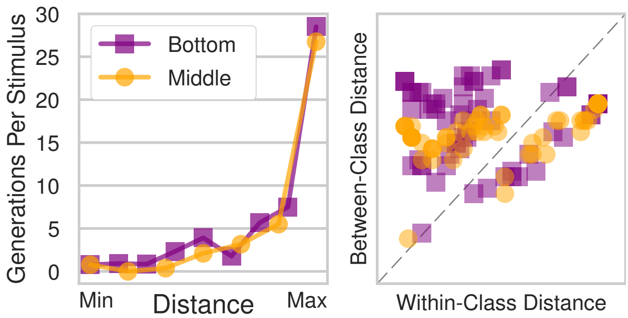

Figure 4: Behavioral results. *Left*: Frequency of exemplar generation as a function of distance from the Alpha category normalized by the maximum possible distance. *Right*: Within- vs. between-category distance for every participant.

different types of categories are generated (with very different distributional information), depending on only the location of the Alphas in the domain.

The results described above are somewhat commonsense: they simply demonstrate that the location (rather than distributional structure) of existing categories imposes constraints on generation because people tend to generate examples in areas not occupied by existing categories. This principle, however, is not predicted by existing models of generation – these models are instead designed to explain distributional correspondences between generated and existing categories.

Model Evaluation

To obtain an overall sense of each model's ability to explain our results, we fit each model by maximizing the log-likelihood of the model's predictions of the human results. The c, ϕ, γ, and θ parameters were fitted for PACKER; c, τ, and θ were fitted for the copy-and-tweak model, and κ, ρ, ν, and θ were fitted for the hierarchical sampling model. Note that each model possess a θ parameter fulfilling the same role (response determinism), and PACKER shares with copy-and-tweak a specificity parameter, c. Attention in PACKER and copy-and-tweak was set uniformly. Parameters were not allowed to vary between participants or conditions – the goal was to obtain the best-fitting values to our entire dataset.

Each model's best-fitting parameterization is shown in Table 2. Overall, PACKER outperformed copy-and-tweak and the hierarchical sampling model by a considerable margin ($\sim$ 17% improvement in log-likelihood). The parameter settings associated with PACKER's best fit are exactly as expected: a strong preference for items that are similar to members of the target category but are dissimilar to members of the contrast category. A similar pattern of results was obtained when we only considered the second to fourth exemplars generated by each participant.

Our model-fitting results follow directly from the design of our experiment. Because the copy-and-tweak and hierarchical sampling models are not influenced by the location of

Table 2: Model-fitting results.

PACKER	**Copy & Tweak**	**Hierarchical Sampling**
$AIC = 3476$	$AIC = 3937$	$AIC = 4061$
$c = 0.565$	$c = 4.675$	$\kappa < 0.001$
$\phi = -4.814$	$\tau = 0.895$	$\nu = 1.001$
$\gamma = 4.250$	$\theta = 5.601$	$\rho = 1.220$
$\theta = 0.731$		$\theta = 7.147$

contrast categories within the space, they do not capture the broad tendency for generated categories to be dissimilar to existing classes.

Individual Differences

Our behavioral results showed that, in general, generated categories consist of items that are dissimilar to members of opposite categories, and similar to members of the their own category. However, beyond this broad description of performance, we observed a great deal of individual differences in generation style. Manually inspecting the data reveals four typical patterns (see Figure 3): 'corners' categories with one Beta example in each corner of the space, tight clusters, 'column'-like categories varying along the y-axis but not the x-axis, and 'row'-like categories varying along the x-axis but not the y-axis. Here, we explore the role of PACKER's attention weights (w_1 and w_2) in producing row-like or column-like categories.

For each stimulus in the domain, we computed the difference in range between the features (e.g., $range(size) - range(color)$) across every generated category in which that stimulus was a member. Taking the average of these range differences yields a gradient describing how, on average, categories were distributed for each stimulus. These data, depicted in Figure 5, reveal a highly systematic relationship between category structure and category location. Whereas column-like categories more often include stimuli to the left or right of the Alpha class, row-like categories appear above and below the Alpha class. Thus, participants modify the distributional structure of new categories to the maximize distance from the contrast category.

To simulate this finding, we set the attention weight parameters in PACKER and copy-and-tweak separately per participant. The other free parameters were set as in Table 2. While there exist methods to find the optimal attention weights for a given classification (see Vanpaemel & Lee, 2012), for ease of computation we simply approximated the weights based on the inverse of each feature's range: $w_k \propto range(k)^{-1}$. Thus, the Alpha and Beta categories are assumed to be distinct along dimensions that the Betas do not vary on. To simulate the hierarchical sampling model we set the domain covariance prior, Σ_0, proportional to the range of each feature (not using the inverse). We then simulated 50 Beta categories with each participant's weighting scheme to obtain a sense of how

Figure 5: Generated category structure as a function of location. Orange areas in each gradient correspond to stimuli that were commonly generated into category possessing greater y-axis range (columns). Purple areas correspond to categories possessing greater x-axis range. White areas correspond to equal range along both features (or infrequent generation).

the relative importance of each dimension affects what types of categories are generated and where they are generated.

The results, depicted in Figure 5, support PACKER's predictions. When the x-axis is weighted more, PACKER creates column categories to the sides of the Alphas. Conversely, when the y-axis is weighted more, PACKER creates row categories above and below the Alphas. This behavior falls out from the nature of selective attention: Dimensions weighted more have a sharper similarity gradient. For example, when the x-axis is weighted more, PACKER favors Beta categories with more within-class similarity (less range), and less between-class similarity along the x-axis, resulting in column-like categories.

Although differentially weighting the features results in different types of categories from the hierarchical sampling and copy-and-tweak models, the location of the Alpha category does not affect where items will be generated by these models. Thus, row- and column-like categories are not systematically generated in different areas of the stimulus space, resulting in the uniform predictions shown in Figure 5.

Discussion

The creative use of conceptual knowledge is a fascinating yet understudied topic in categorization. In this paper, we presented a novel exemplar-based approach to explaining category generation. The PACKER model proposes that categories are represented as a collection of exemplars stored in memory, and that members of generated categories should be similar to one another, yet dissimilar to members of opposing categories. Given that exemplar models can be viewed as Importance-Sampling approximations of Bayesian models (Shi, Griffiths, Feldman, & Sanborn, 2010), PACKER can be viewed as a rational process model and approximating the expected density of a new category based on a contrast category.

In a behavioral study and subsequent formal modeling, we found broad support for the PACKER model. Participants in our study more frequently generated items that are distant from members of contrast categories, and they tended to generate categories with more within-class than between-class similarity. Likewise, we found that the *location* of contrast categories shapes generation by imposing constraints on the areas of space that remain available for a new category. These effects are somewhat commonsense in that they simply reveal that people tend to generate items that are different from what already exists. However, existing models (see Jern & Kemp, 2013) assume little about the role of contrast in category generation (in these models, the location of generated categories is randomly chosen). Our formal simulations reveal that PACKER provides a strong account of the phenomena observed behaviorally.

The PACKER model is, in general, highly expressive in its performance. Under different parameter settings it is capable of generating tightly clustered or highly distributed categories, and adjusting the distribution of categories along each feature. Future work will focus on exploring how individual differences in generation can be explained using PACKER, and whether the canonical effects observed by Ward (1994) and Jern and Kemp (2013) can be explained through exemplar-similarity alone.

Acknowledgments

Support for this research was provided by the Office of the VCRGE at the UW - Madison with funding from the WARF. We thank Kenneth Kurtz for helpful comments, and Alan Jern and Charles Kemp for providing code and data.

References

Jern, A., & Kemp, C. (2013). A probabilistic account of exemplar and category generation. *Cognitive Psychology*, *66*(1), 85–125.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of mathematical psychology*, *15*(3), 215–233.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104.

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, *17*(4), 443-464.

Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, *21*(6), 837–845.

Vanpaemel, W., & Lee, M. D. (2012). Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review*, *19*(6), 1047–1056.

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, *27*(1), 1–40.

Ward, T. B. (1995). Whats old about new ideas. *The creative cognition approach*, 157–178.

Ward, T. B., Patterson, M. J., Sifonis, C. M., Dodds, R. A., & Saunders, K. N. (2002). The role of graded category structure in imaginative thought. *Memory & Cognition*, *30*(2), 199–216.