Creating Something Different: Similarity and Contrast in Concept Generation.

Nolan Conaway[1], Kenneth J. Kurtz[2], & Joseph L. Austerweil[1]

[1]University of Wisconsin-Madison, Department of Psychology, Madison, WI, USA
[2]Department of Psychology, Binghamton University, Binghamton, NY, USA

Author Note
Correspondence concerning this article should be addressed to: Joseph Austerweil, 1202
West Johnson Street, Madison, WI 53706. E-mail: austerweil@wisc.edu

1 Abstract

JLA: Joe! Did you know i made a command for you to enter in comments like this??

```
\jlanote{Some note}
```

Keywords: categorization, concepts, creativity, generation

2 Introduction

The creation of novel concepts and ideas is a highly intriguing – yet infrequently studied – topic of research in human cognition. The creative use of conceptual knowledge is, for example, a core element of scientific investigation, wherein the generation of new ideas is critical to designing experiments and explaining observations. However, due to its complexity, little is presently known about the cognitive processes underlying the generation of new concepts.

Much of what we know about concept generation comes from the foundational literature on creative cognition. In a classic series of reports, Ward & colleagues (Marsh, Ward, & Landau, 1999; Smith, Ward, & Schumacher, 1993; Ward, 1994, 1995; Ward, Patterson, Sifonis, Dodds, & Saunders, 2002) established that category generation is highly constrained by prior knowledge: generated categories tend to consist of features observed in known categories, and they tend to exhibit the distributional properties as found in known categories. In a classic study, Ward (1994) asked participants to generate new species of alien animals by drawing and describing members of the species. People tended to generate species with the same features as on Earth (e.g., eyes, legs, wings), and possessing the same feature correlations as on Earth (e.g., feathers co-occur with wings). Likewise, aliens drawn from the same species tended to share more features with one another compared to members of opposite species.

Much of the work from this area (e.g., Marsh et al., 1999; Smith et al., 1993) focuses on how sample cues (such as an example of a species generated by other participants) can drastically diminish creativity. However, the broader set of observations made by Ward & colleagues provide a great deal of insight into the nature of creative generation. They indicate that people rely strongly on prior knowledge in the creative process, and people generate concepts in accord with what they already know.

Theoretical accounts of these effects have primarily been grounded within the categorization literature. For example, the predominant "Path of Least Resistance"

account (see Ward, 1994, 1995; Ward et al., 2002) proposes that, when generating a new species of animal, people retrieve from memory a known subcategory of animals (e.g., *bird*, *dog*, *horse*), and simply change some of the features to make something new. People were thought to change only features that are not characteristic of the retrieved category (e.g., if *bird* was retrieved, the presence of *wings* would not change, but *color* might). This theory incorporates elements of the highly influential basic-level categories framework (Rosch, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), as well as the exemplar view (Medin & Schaffer, 1978; Nosofsky, 1984, 1986). Problematically, however, the experimental paradigms employed in these studies were relatively uncontrolled, precluding the development of formal approaches. Specifically, participants in these studies are typically asked to generate objects that they have a rich degree of prior knowledge about (e.g., plants, animals, tools, toys), and so it is difficult to apply formal models to the data.

Jern and Kemp (2013) recently showed that creative generation could be studied in a more controlled manner through the well-developed methods of an artificial categorization paradigm (see Kurtz, 2015). In their experiments 3 and 4, participants were exposed to members of experimenter-defined categories of "crystals" varying in size, hue, and saturation. Following a training phase during which the experimenter-defined categories were learned, participants were asked to generate novel categories of crystals. In a finding mirroring that of the Ward (1994) studies, Jern and Kemp found that participants generated categories with the same distributional properties as the experimenter-defined categories: for example, after training on categories with a positive correlation between the size and saturation features (larger sized crystals were more saturated), participants generated novel categories with the same positive correlation. This finding is notable, as it demonstrates that category generation can be studied in a well-known and highly controlled experimental paradigm.

The authors evaluated the predictions of several formal models on their data. Most notably, they showed that a Bayesian hierarchical sampling model provided the strongest

account. Their model represents categories as multivariate normal distributions within the space, and from exposure to category members the model induces a domain-wide representation of the distributional commonalities among the observed categories. When generating a new category, the model samples a new distribution from the domain, and thus the distributional properties of observed categories will be preserved within the generates category.

NBC: Maybe we need a more technical view of the model here?

Jern and Kemp (2013) additionally tested a "copy-and-tweak" model that broadly resembles the earlier "Path of Least Resistance" account: the core proposal is that participants generate new items by copying stored examples from memory and tweaking them to generate something new. The copy-and-tweak model differs from the path of least resistance account in that it notably omits the hierarchical organization of categories, as well as selectivity in which features are changed. Instead, their copy-and-tweak model corresponds to a direct exemplar-similarity approach (e.g., Nosofsky, 1984): participants are thought to generate new items according to their similarity to known members of the target category. This model provided a poor account of the observed data: indeed, the experiments devised by Jern and Kemp were specifically designed to challenge this model. However, it's application is notable as a first step toward explaining category generation using well-known formal approaches from the categorization literature.

2.1   Something Different: A Role For Contrast

It is worth noting that the literature reviewed above provides a somewhat limited view of category generation: the main concern of the published experiments has been on the distributional correspondences between learned and generated categories, and as a result most of the modeling efforts have been towards explaining those effects. In this paper, we investigate another important constraint on the creative use of conceptual knowledge: category contrast. In order to generate a novel concept, individuals must produce

something that is in some capacity *different* from what they already know. Thus, in a trivial sense, contrast can be viewed as a fundamental constraint on creative generation: new concepts must firstly be different from existing ones.

Although it is evident that people are *capable* of creating new concepts and categories, it is not entirely clear how new concepts are systematically made different from what is already known. The hierarchical sampling model developed by Jern and Kemp (2013), for example, does not provide any strong claims about how generated categories should contrast with learned ones: the model only proposes that generated categories are sampled from the same underlying domain distribution as observed categories, and will thus share a common distributional structure. The authors do not make predictions about the *location* of the category within the domain (the perceptual instantiation of category members). Indeed, under a strict interpretation of the hierarchical Bayesian approach, the generated category should actually overlap substantially with the known categories (as it's location is drawn based on those categories). Alternatively, the model's placement of the new category may simply be random (this was the case in the simulations conducted by Jern and Kemp).

NBC: Maybe tune that language to be more in line with the probabilistic world?

The copy-and-tweak model tested by Jern and Kemp (2013) also claims little about how generated categories should contrast with what is already known. In their simulations, the model was only tested on generation after the learner had been exposed to members of the target category, and so the model's ability to generate a new category was not evaluated. However, the model's generation is based exclusively on similarity to known members of the target category; when there are no members of the target category, generation is presumably random.

NBC: Maybe mention something about path of least resistance prediction that contrast is on non-characteristic dimensions?

2.2 Overview

As noted in the sections above, research on the creative generation of novel concepts has thusfar focused on the finding that generated categories tend to possess distributional commonalities with known categories. However, a fundamental goal of concept generation is to create something *new* (i.e., different from what is already known). The manner in which generated categories differ from known ones is, nonetheless, poorly understood: existing formal approaches do not make strong predictions about how creatively generated concepts should systematically differ from existing ones.

In the sections below, we present two experiments demonstrating systematic effects of category contrast on creative generation. Our experiments conform strictly to the artificial category learning paradigm from Jern and Kemp (2013): participants are first exposed to a single, experimenter-defined category, and are then asked to generate members of a new category. We also introduce a novel, exemplar-based model to explain these effects: PACKER (*Producing Alike and Contrasting Knowledge using Exemplar Representations*). Finally, we report formal simulations comparing PACKER to the hierarchical sampling and copy-and-tweak models developed by Jern and Kemp (2013).

3 Experiment 1

To begin our investigation, we developed a artificial, two dimensional domain of squares, varying in color and size (see Figure 2 for samples). We created three unique category types within this domain, shown in Figure 1. The category types were designed to possess strong differences in the distributional structure of the experimenter-defined ('A') category, allowing us to observe the effect of category contrast under a variety of types of prior knowledge. These conditions also provide an opportunity to replicate classic effects observed in previous reports (e.g., Jern & Kemp, 2013; Marsh et al., 1999; Smith et al., 1993; Ward, 1994; Ward et al., 2002).
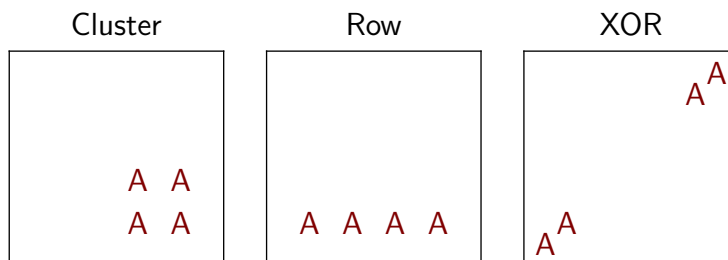
Figure 1: Category types tested in Experiment 1.

> **NBC**: This feels a little weak in motivation: how do these conditions specifically allow us to show is these effects? Did we come up with these out of thin air? So maybe foreshadow that you hardly need to look for these effects, they're *everywhere*. Maybe there is also a problem that at this point it is not exactly clear what contrast means within this scenario.

In the 'Cluster' type, the experimenter-defined category is a tight cluster of examples in the space (e.g., all members of category 'A' are small and dark in color). In the 'Row' type, the category has a row pattern across the space: varying along one feature but not the other (e.g., members of category 'A' are small and vary in color). In the 'XOR' type, the experiment-defined category consists of two clusters separated in opposite corners of the space category, conforming to the exclusive-or logical structure (e.g., members of category 'A' can be small and dark or large and light).

3.1   Participants & Materials

183 participants were recruited from Amazon Mechanical Turk. Participants were randomly assigned to one condition: 64 participants were assigned to the Cluster condition, 61 were assigned to the Row condition, and 58 were assigned to the XOR condition. Stimuli were squares varying in color (RGB 25–230) and side length (3.0–5.8cm). These stimuli are slight variants of those used by Conaway and Kurtz (2016), see Figure 2 for samples. The assignment of perceptual features (color, size) to axes of the domain space
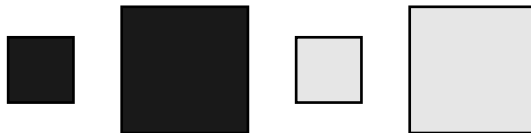
Figure 2: Sample stimuli used in Experiments 1 and 2.

(x, y) was counterbalanced across participants.

3.2 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. Each block consisted of a single presentation of each of the members of the Alpha category, in a random order. 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'. As in Jern and Kemp (2013), generation was completed using a sliding-scale interface. Two scales controlled the features (color, size) 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. Prior to beginning the generation phase, participants read the following instructions:

> As it turns out, there is another category of geometric figures called "Beta".
> Instead of showing you examples of the Beta category, we would like to know
> what you think is likely to be in the Beta category.

You will now be given the chance to create examples of any size or color in order to show what you expect about the Beta category. You will be asked to produce 4 Beta examples - they can be quite similar or quite different to each other, depending on what you think makes the most sense for the category.

Each example needs to be unique, but the computer will let you know if you accidentally create a repeat.

Following generation, participants completed a generalization phase wherein they classified novel examples into the Alpha and Beta categories without feedback. On each trial, a single example was presented, and participants were asked to classify it by clicking buttons labeled "Alpha" or "Beta". Participants classified a total of 81 items sampled along a 9x9 grid, including the members of the Alpha and Beta categories (randomly intermixed). These data were, however, collected to address a separate set of questions, and we do not discuss them in this paper.

4   Experiment 2

4.1   Participants & Materials

122 participants were recruited from Amazon Mechanical Turk. 61 participants were randomly assigned to the Middle and Bottom conditions each. The stimuli were exactly as in Experiment 1. Again, the assignment of perceptual features (color, size) to axes of the domain space (x, y) was counterbalanced across participants.

4.2   Procedure

The procedure was exactly as in Experiment 1: participants first completed a short training phase, followed by the generation phase, followed by the generalization phase (data from this phase is not discussed in this report).

5 The PACKER Model

The PACKER model is an extension of the influential Generalized Context Model of category learning (GCM; Nosofsky, 1984, 1986). The GCM simulates categorization under the assumption that learners represent categories as a collection of exemplars, corresponding to the labeled stimuli they have observed. The exemplars are encoded within a k-dimensional psychological space, and categorization is based on the amount of similarity between the item to be categorized and the stored exemplars. Similarity between two examples, $s(x_i, x_j)$, is computed as an inverse exponential function of distance (following Shepard, 1957, 1987):

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

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 in similarity calculations, and c ($c > 0$) is a specificity parameter controlling the spread of exemplar generalization. For simplicity, attention will be distributed uniformly in our simulations (unless otherwise noted). The value of r depends on the nature of the experimental conditions being simulated: $r = 1$ is appropriate for separable dimensions, whereas $r = 2$ is appropriate for integral dimensions (see Garner, 1974; Shepard, 1964).

PACKER's core proposal (as well as its name) was in part inspired by earlier work from category learning literature (see Hidaka & Smith, 2011). PACKER proposes 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: the most desirable generation candidates are similar to members of the target category and not similar to members of contrast categories. This is achieved by aggregating similarity across known exemplars differently according to class membership. The aggregated similarity a between generation candidate y and stored exemplars x is given by:

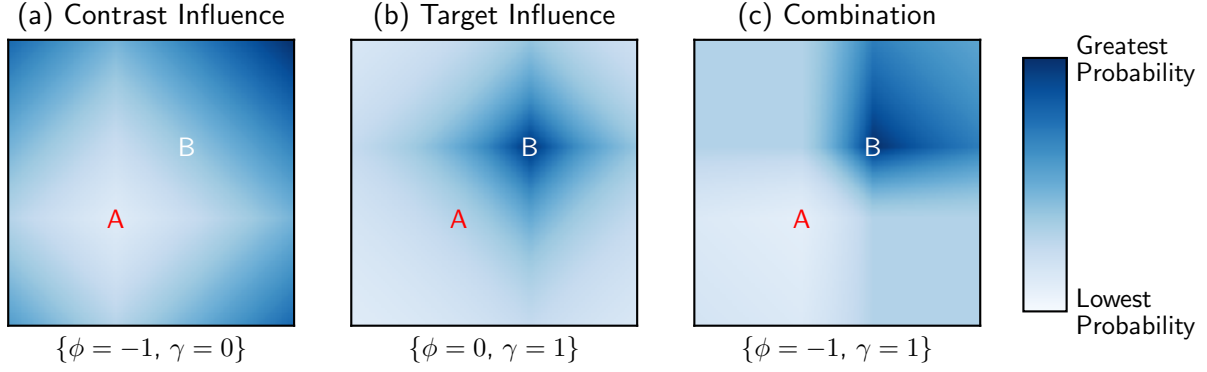| (a) Contrast Influence	(b) Target Influence	(c) Combination
$\{\phi = -1, \gamma = 0\}$	$\{\phi = 0, \gamma = 1\}$	$\{\phi = -1, \gamma = 1\}$

Figure 3: PACKER generation of a category 'B' example, following exposure to one member of category 'A' and one member of category 'B'. Predictions are shown for three different parameterizations: *(a)* Predictions based on contrast similarity only. *(b)* Predictions based on target similarity only. *(c)* Predictions with both constraints considered.

$$a(y, x) = \sum_j f(x_j)s(y, x_j) \tag{2}$$

where $f(x_j)$ is a function specifying each exemplar's contribution to generation. PACKER sets $f(x_j)$ depending on exemplar x_j's category membership: $f(x_j) = \phi$ if x_j is a member of a contrast category, and $f(x_j) = \gamma$ 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 result in greater consideration of those exemplars, with values of 0 eliminating their effect. A negative value for $f(x_j)$ produces a 'repelling' effect (exemplars are less likely to be generated nearby x_j). Conversely, a positive value for $f(x_j)$ produces an 'attracting' effect (exemplars are more likely to be generated nearby x_j).

As noted above, PACKER proposes that new categories should be different from existing categories, and same-category exemplars should be similar to one another. This is realized when $\phi \leq 0$, and $\gamma \geq 0$. Negative ϕ values encourage y to be distant from contrast categories (as similarity to contrast category exemplars are subtracted during aggregation). Positive γ values encourage y to be close to other exemplars of the target category. When

$|\phi| = \gamma$, the repulsion effect from contrast categories is equal to the attraction effect to the target category. See Figure 3 for an illustration of how these parameters are used.

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:

$$p(y) = \frac{\exp\left\{\theta \cdot a\left(y, x\right)\right\}}{\sum_i \exp\left\{\theta \cdot a\left(y_i, x\right)\right\}} \tag{3}$$

where θ ($\theta \geq 0$) is a free parameter controlling response determinism.

NBC: It is worth noting that PACKER is not the only *possible* model that could incorporate a contrast mechanism. However, contrast falls out naturally from the exemplar view and ... Contrast could, for example, be implemented within...

5.1 Relation Between PACKER and Copy-And-Tweak

The PACKER model is in many senses similar to the copy-and-tweak model reported by Jern and Kemp (2013): both models are exemplar-based, and both models generate new items according to their similarity to known members of the target class. PACKER diverges from from the copy-and-tweak model only in the inclusion of a contrast mechanism, enabling generation according to dissimilarity to members of opposing categories. By consequence, copy-and-tweak can be realized as a parameterization of the PACKER model that is insensitive to category contrast. Specifically, when $\phi = 0$ and $\gamma = 1$ (see Figure 3, panel B), PACKER is not influenced by similarity to members of opposite categories, and is mathematically equivalent to a copy-and-tweak approach.

In the sections below, we report simulations using this copy-and-tweak model. Due to their mathematical equivalence, the comparison between PACKER and copy-and-tweak provides a test of the explanatory value of the contrast mechanism: the account provided by copy-and-tweak will only equal that of PACKER if the contrast mechanism does not

offer an advantage.

5.2 The Hierarchical Sampling Model

Jern and Kemp (2013) demonstrated how a hierarchical Bayesian model could explain the distributional correspondences between observed and generated categories. In their model, exemplars of generated category were viewed as samples from a multivariate Normal distribution over the dimensions of stimulus space. The mean of the generated category was independent of the observed categories, but the covariance matrix (encoding feature variances and correlations) was based on a common prior distribution. Generating a new category was thus completed by sampling a new category mean (uniform over stimulus space) and covariance matrix from the common prior distribution. Because the shared prior distribution's parameters were unobserved, the hierarchical Bayesian approach was used to infer its parameters from the previous categories (their feature variances and correlations), and then to generate the covariance matrix of the new category.

In our implementation of their model, each category's exemplars are assumed to be sampled from a multivariate Normal distribution with parameters (μ, Σ). Each category's covariance matrix is assumed to be inverse-Wishart distributed with parameters (v, κ, and Σ_D).[1] Σ_D is the covariance matrix shared between categories. We assume the shared covariance matrix Σ_D is generated from a Wishart distribution (for conjugacy) with parameters v_0, κ_0, and Σ_0. We set $\nu_0 = 4$, and $\Sigma_0 = \rho\mathbf{I}$, where ρ is 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.

NBC: check on that $\nu_0 = 4$ business

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 $\Sigma_D = \Sigma_0 + \sum_c C_c$,

---

[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.

where C_c is the empirical covariance matrix of category c. The MAP estimate of the covariance matrix for the target category B is

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

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.

Generated exemplars are drawn from a multivariate Normal distribution specified by (μ_B, Σ_B). Thus, $p(y)$ is

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

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

6   Acknowledgments

References

Conaway, N., & Kurtz, K. J. (2016). Similar to the category, but not the exemplars: A study of generalization. *Psychonomic Bulletin & Review*, 1–12. doi: 10.3758/s13423-016-1208-1

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Hidaka, S., & Smith, L. B. (2011). Packing: a geometric analysis of feature selection and category formation. *Cognitive Systems Research*, *12*(1), 1–18.

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

Kurtz, K. J. (2015). Human category learning: Toward a broader explanatory account. *Psychology of Learning and Motivation*, *63*, 77–114.

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

Marsh, R. L., Ward, T. B., & Landau, J. D. (1999). The inadvertent use of prior knowledge in a generative cognitive task. *Memory & Cognition*, *27*(1), 94–105.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207.

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

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325–345.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87.

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

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.

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

Ward, T. B. (1995). What's old about new ideas. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 157–178). Cambridge, MA: MIT Press.

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.