# Gradient similarity within compositional representations

Lauren A. Oey

Department of Brain and Cognitive Sciences, University of Rochester

A Senior Honors Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Honors in Bachelor of Science

April 2018

Supervised by Steven T. Piantadosi

Abstract

When learning about the world, we develop mental representations or concepts for things that can be defined using overarching rules, known as rule-based systems. At the same time, we also develop representations for things that are similar to what we have experienced, known as similarity-based systems. Traditionally, rule-based and similarity-based systems have been used as distinct models to capture conceptual representation. However, it seems implausible that we do not flexibly deploy both systems. Whether both systems can be used simultaneously to represent components of a single concept is an open empirical question. One example suggesting that the use of both systems is possible is the concept of a ZEBRA, which *looks like a horse but striped*. In this description, *looks like a horse* relies on similarity and *striped* relies on rules. To address our question, in Experiment 1, we use an artificial concept learning task to test whether people can combine similarity and rules compositionally in order to represent concepts with one Boolean and one continuous dimension. Our results suggest that participants are not only able to compose similarity and rules, but they are also able to retain a gradient representation of the similarity-evaluated dimension. If gradient information can be retained within a compositional system, this opens the question of how people evaluate the conjunction of components across two different similarity-based dimensions. We test this in Experiment 2 using a similar task to Experiment 1, except with stimuli having two continuous dimensions. To infer what computations people are using to evaluate combinations of both similarity- and rule-based components, we use the results from Experiment 2 and fit multiple proposed models that are qualitatively consistent with the expected results in concepts with either two Boolean dimensions, one rule-based and one similarity-based dimension, or two continuous dimensions. Our results show that some models (e.g., minimum-value, Euclidean distance, weighted averaging) better represent mental representations of conjunctions over similarity-based components than others (e.g., maximum-value, hard-threshold). The better-fitting models each demonstrate the intuitive characteristic that some features are more relevant to our concepts than others.

*Keywords:* concept learning, conceptual representation, compositional systems, similarity, rules, generalization

# Gradient similarity within compositional representations

Concepts are at the core of how humans develop an understanding of the world around them. That being said, exactly how concepts are represented in the human mind has long been debated within the field of cognitive science (Margolis & Laurence, 1999). Traditionally, at least two distinct systems have been used to describe conceptual representation (Hahn & Chater, 1998): similarity (e.g., Shepard, 1980; Newell & Simon, 2007) and rules (e.g., Nosofsky, 1984). Research in the conceptual representation domain often attempts to construct a core distinction between these two systems. Given the assumption that we use both similarity- and rule-based representational systems, it still remains an open question as to how the systems interface within a single concept. How do we represent concepts such as ZEBRA, which can easily be and is readily described as being *like a horse but striped*, where the concept seems to require both similarity- and rule-based components?

Both similarity- and rule-based systems have much to offer in terms of explanatory power. Similarity-based systems rely heavily on memory for perceptual features. In similarity-based systems, previously seen examples are sometimes represented as noisy feature vector. This format has two distinct advantages. First, features do not need to co-occur deterministically in order for some abstraction to be made. As a result, characteristic features of concepts are stored and used to aid recognition. Second, similarity-based systems flexibly handle partial matching during generalization tasks—i.e. a novel item does not need to strictly match the abstract representation of a given category on all dimensions in order to be grouped into that same category. This is a desirable feature considering the rampant inconsistencies in our world (e.g., most but not all chairs have legs).

Alternatively, rule-based systems typically handle discrete feature values. Rules of this system naturally compose to generate increasingly complex but precise rules (Feldman, 2000). Rules require less memory storage, as people only need to store the rule information rather than instances within the category. Additionally, rules are easily verbalized, making them prime for linguistic transmission. Although rules may act as a useful tool in defining categories, they do not handle fuzzy, non-discrete values well, and the environment is filled with naturally continuous features, e.g. color. Fortunately, linguistic labels have been shown

to affect how people divide the mental space into categories. Language facilitates how information is stored by highlighting perceptual distinctions in the world and serving as a cue to attend to that distinction and form a category (Waxman & Markow, 1995). In color space, it has been shown that a distinction in blue color labels that is present in Russian but not in English contributes to differences in perceptual color discrimination tasks (Winawer et al., 2007). These findings provide evidence that language can affect perceptual categorization, and more generally, facilitates the discretization of naturally continuous spaces.

Research on cognitive development suggests that the use of these two systems may transform over time when acquiring a given concept (Werner, 1948; Bruner, Olver, & Greenfield, 1966; Kemler, 1983). Children initially learn using a similarity-based system, but later develop a rule-based representation. The early similarity-based system allows a child to make a judgment without explicitly knowing the relevant features (Kemler, 1983). For example, a child can judge holistically whether an object is a diamond based on its similar appearance to previous diamonds seen. The child does not need to understand the concept of an equilateral polygon (i.e. all sides on a figure being equal) in order to make this judgment. However, after learning about equilateral polygons, a child will be able to make better generalization judgments on novels objects that are not visually similar to prototypical diamonds.

Some hybrid theories exist including RULEX (Nosofsky, Palmeri, & McKinley, 1994). In RULEX, the model learns rules and exceptions to those rules. While similarity is not explicitly implicated, the flexibility provided by the exception allows for the model to capture similarity based generalizations. Additionally, probabilistic rule-based systems capture a more expansive set of conceptual components. In Goodman, Tenenbaum, Feldman, and Griffiths (2008)'s rational rules model, a probability distribution over rules is represented, which allows similarity to be captured by the probability gradient over rules. One other model that more explicitly includes similarity and rules is Heit and Hayes (2011)'s GEN-EX model, which incorporates rules into the notion of similarity embodied by the Generalized Context Model (Nosofsky, 1988) to explain inductive reasoning behavior.

Although these hybrid theories begin to address how similarity and rules can be

integrated into a single model, previous research still does not specifically address how similarity and rules play a role at the feature level to inform our overall representation. In the current literature, it is experimentally unclear whether the evaluation of gradient features can be incorporated in rule-based models. If the evaluation of gradient features can be integrated into a compositional system as a component, will this component capture patterns of similarity, such as the preservation of continuity within the representation.

Our goal in this paper is to see how people freely combine gradient and discrete judgments by combining similarity components into larger compositional systems. We argue that both similarity and rules are necessary and can be combined to form every representation. We conducted an experiment to examine whether similarity- and rule-based systems can compose to represent a single concept with both continuous and Boolean features. We found evidence for the compositionality of both similarity and rules. If similarity can act as a component within compositional systems, this opens the question as to how components are evaluated in combination with each other, i.e. two rule components, one rule and one similarity component, and two similarity components. Given that the generalization results of conjunctions over two rule components are clear from the logic literature, and our first experimental findings provide evidence for the case of conjunctions over rules and similarity components, we then conducted an experiment to examine how two similarity-based components are integrated into a single concept.

## Experiment 1

To investigate whether learners can compose similarity- and rule-based mental systems, we used an artificial concept learning paradigm. Participants were asked to make generalization judgments (i.e. whether the label applied to a new item) for a novel category of objects, *feps*. Critically the concept being learned contained two relevant features: one that is continuous—i.e. not readily discretized into categories or easily described with language, and one that is Boolean—i.e. taking discrete categories and easily described with language. On the continuous dimension, *feps* were blob-y shaped; and on the Boolean dimension, *feps* were always filled in white, as opposed to black. One assumption we made was that the continuous

feature would elicit the use of similarity-based representations and the Boolean feature would elicit the use of rule-based representations. We hypothesized that if participants represent a concept with both continuous and Boolean features, people would generalize along a system that incorporated both features. If participants retain a gradient representation in concepts combining both features, we can conclude that people can combine similarity- and rule-based representations compositionally and flexibly. However, participants may not necessarily retain a gradient representation, instead discretizing the continuous space (i.e. a continuous feature is "similar-enough"), and this would suggest that participants are not maintaining the gradient similarity information within the compositional system. Alternatively, participants may only generalize along either the continuous or Boolean dimension, suggesting that participants use each system separately.

## Participants

We recruited 106 participants on Amazon Mechanical Turk. Three participants were excluded because of their failure to complete the task. Two additional participants completed the non-linguistic task, but failed to complete the linguistic task. These participants' non-linguistic data was included in the analysis.

## Stimuli

The stimuli consisted of 100 unique images, each of which was manipulated along two features: shape and fill color. Critically, we manipulated the features such that one would be continuous (i.e., shape) and the other would be Boolean (i.e., fill color). Along the shape dimension, there were 50 shapes. Shapes were generated using a custom python script and were outputted to an SVG file[1]. The curvature of the arcs were determined by a normal distribution and the coordinates of each of the four vertices were determined by independent uniform distributions. In other words, along both the x- and y-axes for each of the four vertices, there were three points marking the extreme edges of two uniform distributions, where one edge was shared among the distributions (Figure 1). This generating function lends

[1] Code avalilable at `github.com/loey18/Oey_Zebra/`

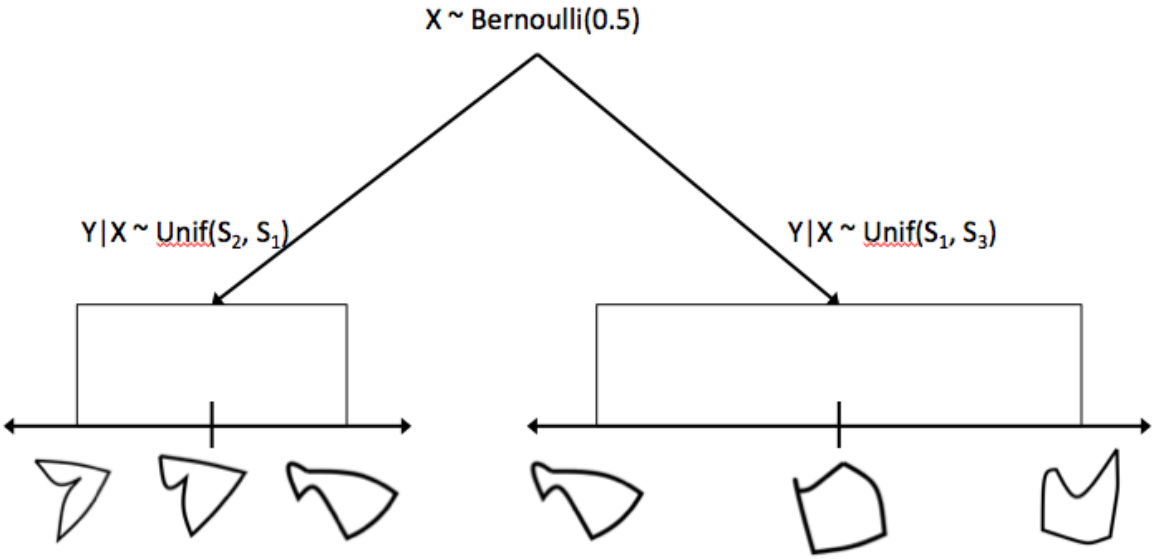


*Figure 1*. Function used to generate the stimuli shape. Bernoulli distribution followed by a uniform distribution is used to determine what shape is generated. Both uniform distributions share the same shape at the "edge value."

itself to forming a prototype, where the shared distribution edge along each vertex acts as this prototype.

Along the fill color dimension, the stimuli were either filled in black or white. This was manipulated by manually adjusting the fill color, creating both a black and white filled image for each shape.

In order to measure how similarly the gradient feature (shape) would be perceived, we collected norming data ($n = 24$). Six participants were excluded from the analysis because greater than 50% of their responses had similarity rating scores that were less than 5 out of 100. We asked participants to adjust a slider in order to rate each shape's similarity to a single shape acting as the reference point. The reference point used was the prototype described above (Figure 1). Importantly, the fill color was held constant across all shapes. We then normalized each of these ratings relative to the given participant's responses and used the means of these normalized z-scores as the similarity measurement for a given shape (Figure 2). The reference shape when rated in comparison to itself is represented by the far right bar and has the highest similarity rating, acting as a sanity check.

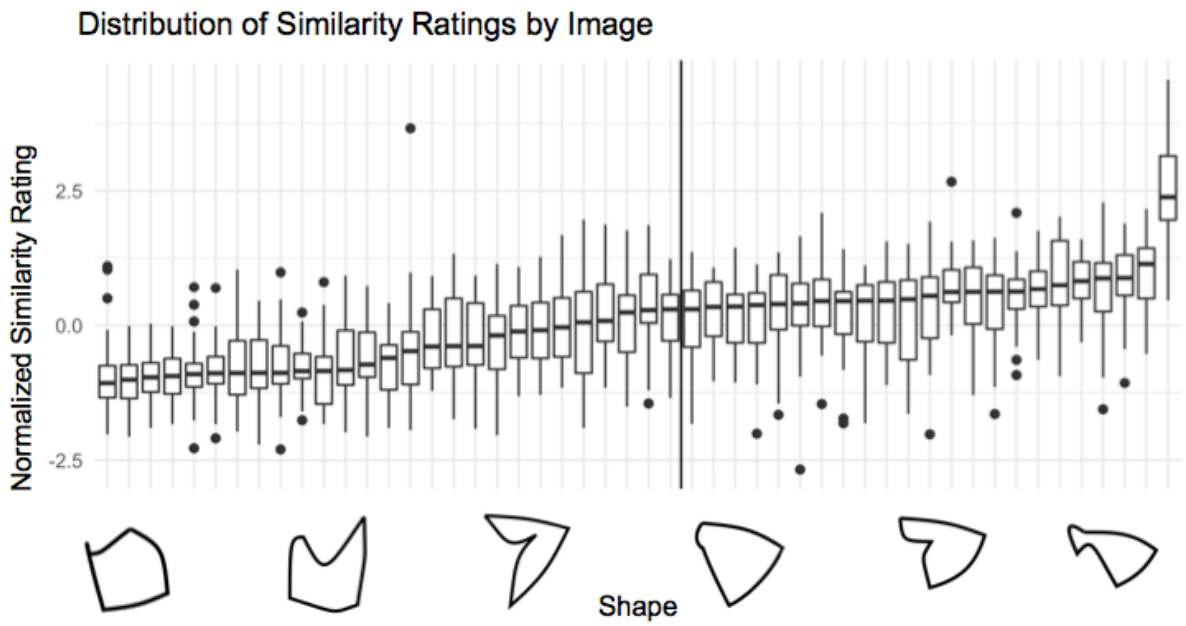


*Figure 2*. Normalized similarity ratings for each of the stimuli shapes. The x-axis indicates individual shapes, and the y-axis represents the similarity rating, normalized by participant. The boxplots show the median and quartiles of normalized ratings for each shape. The upward slope in the median normalized ratings across items express that the similarity of the shapes is perceived gradiently. The vertical line indicates the cutoff point used for determining which shapes would be included within the set of *feps* in the concept learning task.

## Procedure

There were two parts to the experiment: a generalization and a verbal description task (Figure 3).

**Generalization Judgment (Non-Linguistic) Task**. Participants were shown exemplars of a novel object *fep*, which is analogous to how we learn about concepts in the real world through exposure to examples. All exemplars were white and similar to the prototype, or the reference shape used to collect norming data.

For each trial, participants were shown a labelled exemplar of a *fep*, which remained on the screen for the remainder of the experiment to reduce memory load. They were then shown ten novel test items that were sampled randomly without replacement from the list of stimuli. Participants received feedback after each generalization judgment, being shown either a green check mark if correct, or a red X if incorrect. The experimenters decided upon a subset of the

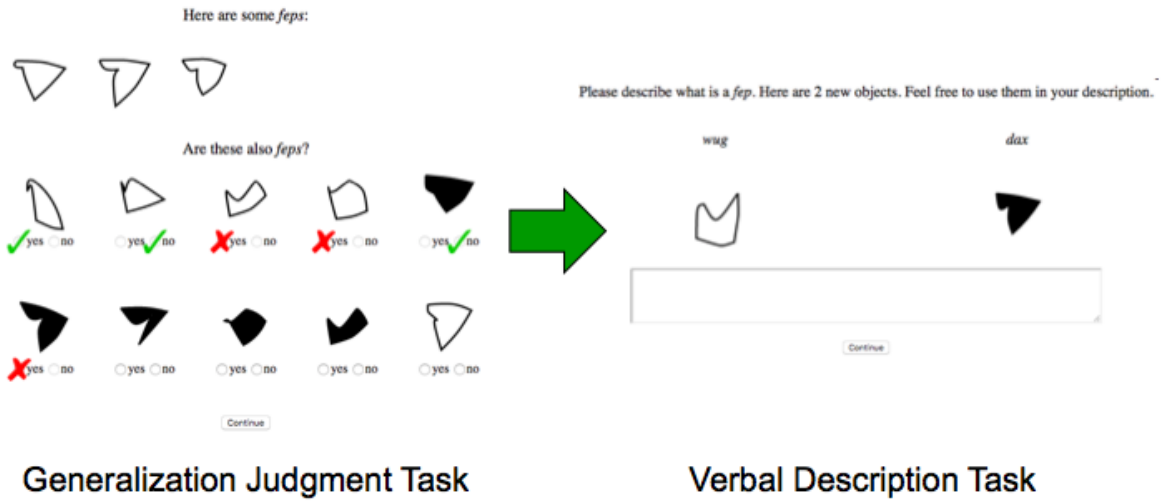


*Figure 3*. The experiment design. The first part of the experiment consisted of a generalization judgment task, followed by a verbal description task.

stimuli that would be labelled as a *fep* when providing feedback. There were 23 items that were deemed as *feps*, and this subset consisted of the 23 shapes that received the highest similarity ratings and were also white filled.

Participants saw a total of six *fep* exemplars incrementally over six trials. The *fep* exemplars were hand-selected by the experimenter and were consistent across all participants. The order in which these exemplars were displayed was randomized for each participant. Participants made a generalization judgment on each of the 60 test items. None of the exemplars acted as a test item. As the total set of stimuli consisted of 100 items, and participants were exposed to the six exemplars and 60 test items, participants only saw a subset of the total stimuli.

**Verbal Description (Linguistic) Task**. At the end of the experiment, participants were asked to provide a description of a *fep*. Participants were provided with two novel labelled images to act as potential modulus in their descriptions, though they were instructed that they were not required to mention these. One of the objects (*wug*) was dissimilar in shape to a *fep* but shared the white fill feature. The other object (*dax*) was filled in black but was similar in shape to a *fep*.

The description task served two purposes. First, the task acted as a sanity check to

verify that participants did not induce the same set of rules defining the boundaries of the *fep* shape space. If participants universally induced some set of rules (e.g. flat top and curved bottom) that would capture the items within the subset of images pre-experimentally labelled as *feps*, this would suggest our manipulation failed. It would also speak to the remarkable human ability to easily induce rules, providing evidence against the default use of a similarity-based system.

Second, this task allowed us to compare the representations suggested by both the linguistic and non-linguistic data. For example, participants may exclusively use rule-type language, when their non-linguistic data reflects more of a similarity-based representation. One could also examine the frequency of modal or gradable language (Lassiter, 2017) used in the descriptions, to distinguish between a similarity function and a probabilistic rule, which may produce similar empirical results.

## Possible Outcomes

**Generalization Judgment (Non-Linguistic) Task.** We primarily considered four potential outcomes from the non-linguistic experimental task:
Participants may only generalize their concept of a *fep* along one of the two critical features, suggesting that people use the representation systems separately.

(a) Participants may exclusively generalize along the Boolean feature (i.e. black versus white fill), suggesting the use of a rule-based representation (Figure 4, top left).

(b) Similarly, participants may exclusively generalize along the gradient feature (i.e. shape), suggesting the use of a similarity-based representation (Figure 4, top right). Alternatively, participants may consider both the Boolean and gradient feature when making generalization judgments.

(c) One potential outcome is that both Boolean and gradient features will be considered when making generalization judgments; however, the gradient feature may in fact be evaluated along a discretized rule (Figure 4, bottom left). Participants may align on some threshold to indicate an item is “similar-enough” to some other item, and items below that threshold are not “similar-enough.” In other words, within such a compositional structure, similarity would

be mapped onto rules.

(d) If the Boolean feature is evaluated using a rule-based system and the gradient feature is evaluated using a similarity-based system, we would predict results similar to the bottom right graph in Figure 4, where the effect of gradient feature will be evaluated differentially, depending on the Boolean feature. In other words, similarity would preserve continuity in conjunction with discrete features within a larger compositional structure.

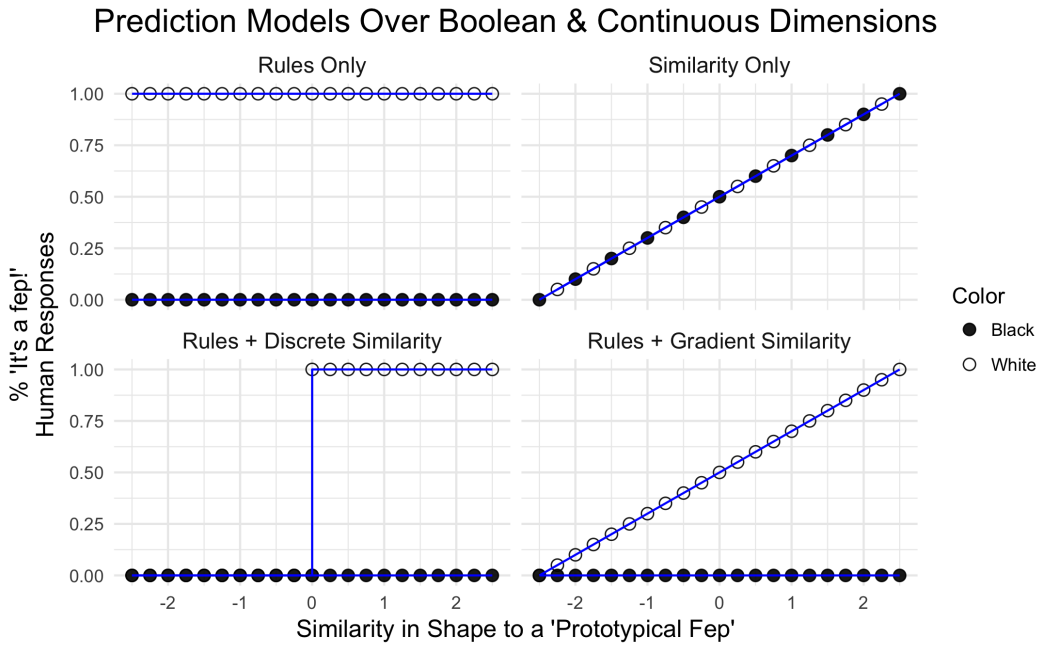


*Figure 4*. Prediction graph for each of the four proposed hypotheses for the representation of concepts with one discrete and one continuous dimension. The x-axis represents similarity on a scale of $-2.5$ to 2.5. The y-axis represents the percent of instances that a given item is generalized to (i.e. a test item is predicted to be a *fep*). The line types represent the items with the different discrete features. [Top left] Participants only generalize along the Boolean feature (fill color); [top right] participants only generalize along the gradient feature (shape); [bottom left] participants generalize along both the Boolean and gradient feature but the gradient feature is evaluated as a discrete feature; and [bottom right] participants generalize along both the Boolean and gradient feature and the gradient feature is evaluated as a continuous feature.

**Results**

Figure 5 shows the experimental results of the generalization judgments. The x-axis represents the normed similarity ratings for each test item shape. Values range from roughly from $-1$ to 2.5. The y-axis represents the proportion of responses that answered "yes" to the question of whether a given test item was also a *fep*, which varied from 0 to 1. The different colors of the data points represent the fill color of the test item. Critically these results closely resemble the predicted results in the bottom right of Figure 4 (d). This suggests that participants are evaluating both the discrete and gradient feature compositionally, and in assessing the gradient feature, a similarity-like continuous representation is preserved.

We analyzed conceptual representation based on the participant generalization judgments on the last three trials. We assumed that participants would develop a representation of the concept in the first half of the experiment (i.e. first three trials), and we will examine the learning aspect separately below.

We used a logistic mixed-effects regression model to analyze the data from the last three trials. The response variable is the generalization judgment coded as yes, generalize = 1 and no, do not generalize = 0. The model examined whether there was an interaction between stimuli similarity rating on shape and the stimuli's fill color. The fill color independent variable was dummy coded, with black fill as the referent level (black = 0, white = 1). Additionally, the model contained by-participant, by-test item, and by-exemplar random effects. The data was analyzed in R using the *glmer* function of the linear model package, *lme4* (Bates, Mächler, Bolker, & Walker, 2014).

We found a significant effect of fill ($\beta = 2.175$, $z = 14.445$, $p < 0.0001$), similarity ($\beta = 0.625$, $z = 3.888$, $p < 0.0002$), and the interaction between similarity and fill ($\beta = 0.904$, $z = 3.713$, $p < 0.0003$) (Figure 5). These results suggest that people are able to combine similarity- and rule-based systems when developing a representation for the concept *fep*. This hypothesis is further reaffirmed by the distribution's similar shape to the prediction graph in the bottom right graph in Figure 4.

In examining the mental representation of the gradient feature, it was important that we assess whether participants evaluate the feature continuously or discretely. Showing that

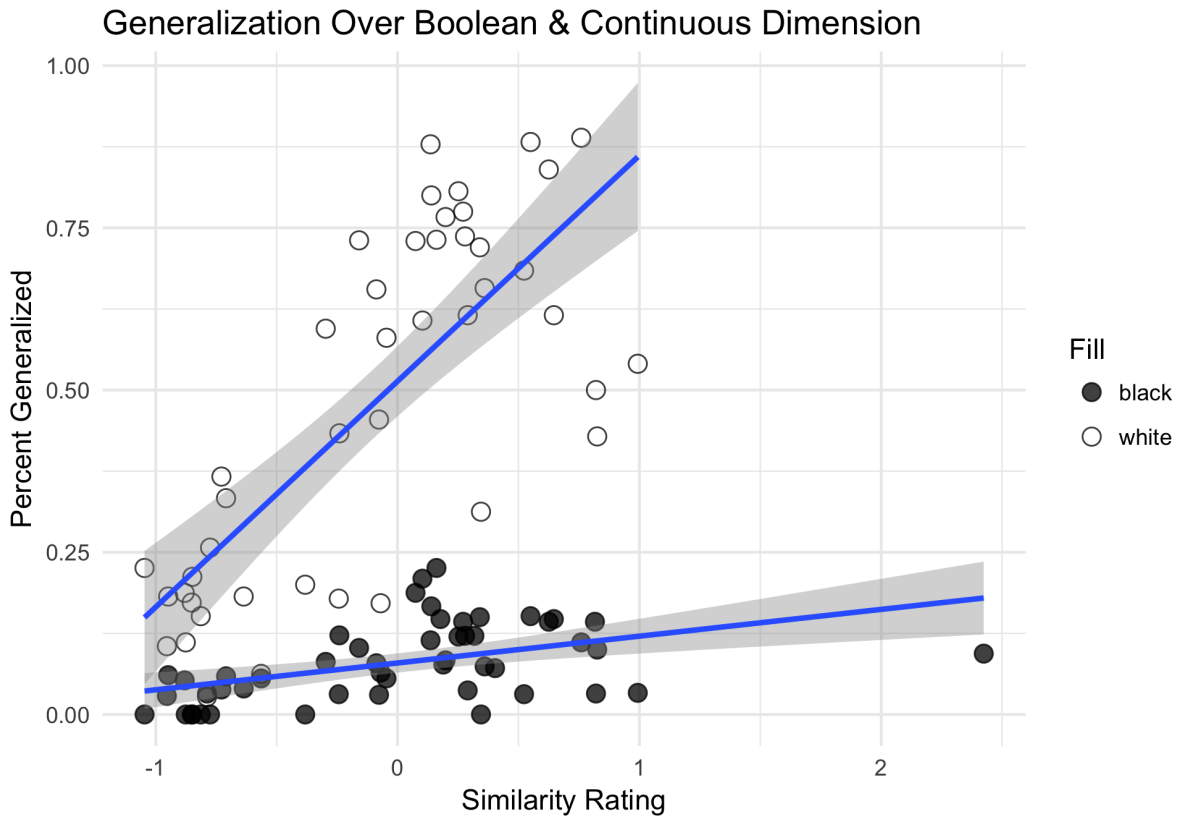


*Figure 5*. Results from Experiment 1 artificial concept learning experiment, in which objects have one Boolean and one continuous feature. There is an interaction between the Boolean feature (object fill color; represented as point color) and perceived similarity (shape; represented on x-axis) affecting generalization judgments (y-axis).

aggregated generalization judgments along the gradient feature are linear provided supporting evidence for a continuous representation. However, we might also see a linear pattern if participants had individually-variable, discrete similarity thresholds. To address this alternate hypothesis, we fit a logistic regression for each participant individually. The model predictions are visualized in Figure 6. We do not find sharp step-wise functions reflecting discrete similarity thresholds. This finding is, however, consistent with flexible, combinatorial deployment of similarity and rule based systems.

Surprisingly, there is a significant effect of similarity, even with only the black filled test items. One potential explanation for this surprising result is that participants are not learning the rule over the fill color feature to the extent that we expected they would. There may still be

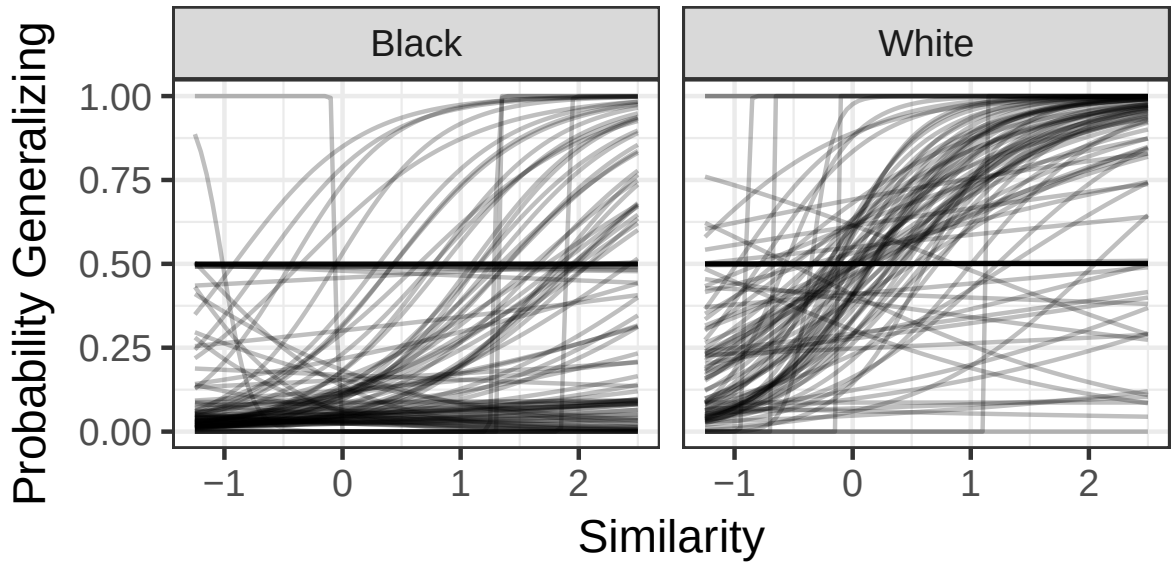


*Figure 6*. Predictions for individuals generalization probability as a function of similarity for black and white stimuli. The gradient slope for similarity among the white images, suggests participants do not have step-like thresholds for similarity.

a belief in the relevance of shape similarity, even when the evidence should suggest that similarity should be irrelevant given the rule. However, it is also important to note that the slope is relatively flat. As the participant pool is large ($n = 103$), we may be detecting small effect sizes.

**Learning**

A visualization of a qualitative pattern of category learning over the course of the experiment's six trials is shown in Figure 7. Each panel corresponds to generalization judgments in a given trial.

Data points in the initial trials are more scattered, but over the course of these first few trials, the data begins to appear more consistent across trials. In our data, by around the third trial, participant data is beginning to show a positive slope for the white filled items and a flatter slope for the black filled items. Over the course of the six trials, the data shows a pattern emerging, where there is a divergence in the lines representing generalization judgments by the discrete feature. This shows that in the initial trials, participants relied heavily on

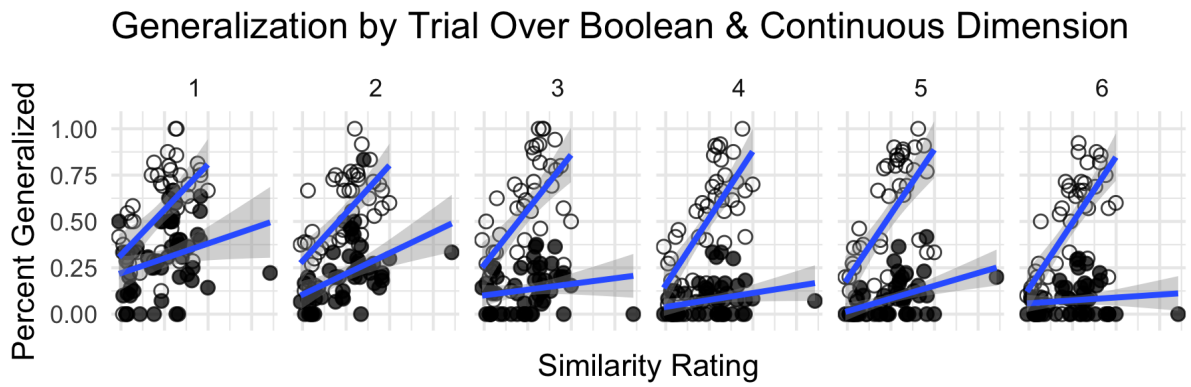


*Figure 7*. Change in representation over the course of six trials in the Experiment 1 artificial concept learning task, in which objects have one Boolean and one continuous feature. There is an interaction between the Boolean feature (object fill color; represented as point color) and perceived similarity (shape; represented on x-axis) affecting generalization judgments (y-axis).

similarity measurements on the continuous feature across both the black- and white-filled items, even though only the white-filled were presented as being *feps*. This would suggest that participants are relying on similarity systems initially, and after gaining more exposure, integrate rules into their concept. Follow-up research should examine how representation unfolds during learning when the relevant features are less predictable.

Initial trials seem to show general positive sloping trends between generalization judgments and similarity ratings across both fill colored items. This is indicative of a shape bias (Landau, Smith, & Jones, 1988). Additionally, the slope of the white filled items are shifted upward relative to the black filled items, indicating a tendency to generalize to items that share the same color as well. These categorization patterns in the initial trials indicate a similarity-based representation, which is consistent with the view that similarity is free, i.e. similarity representations are quickly learned.

**Language**

We predicted that participants would describe a *fep* using language of both rules and similarity. One such predicted verbal response would be "A *fep* is white and like a *dax*." However, participants may exclusively use similarity-like language (e.g. "A *fep* is like a *dax* in

shape and a *wug* in color.")

The verbal responses were individually hand-coded by the first two authors into mutually exclusive categories (i.e. yes, feature is present vs. no, feature is not present) along two dimensions (i.e. similarity-like and rule-like language). The inter-rater agreement was measured at a Cohen's kappa coefficient of 0.678.[2] The discrepancies were debated, and a post-reconciliation kappa coefficient was found to be 0.989. The results of this classification are presented in Table 1.

| | $\neg Similarity$ | $Similarity$ | **Total** |
|---|---|---|---|
| $\neg Rules$ | 8 | 6 | 14 |
| $Rules$ | 50 | 37 | 87 |
| **Total** | 58 | 43 | 101 |

Table 1

*Counts of the 101 participant verbal descriptions, categorized by uses of similarity and rule representations in language.*

As predicted, we find that people do in fact use both rules and similarity-like language, and this occurred in 36.6% of responses (e.g. "A *fep* is more like a *dax* but white."). It is also important to consider that 86.1% of participants using rules in their language, suggesting that rule-based representations are preferred in language. A key contributing factor is that language is often considered to be rule-like; thus, it may be unsurprising that the majority of participants use rule-like language in their descriptions.

There is one prominent limitation to our verbal description task: including the moduli (i.e. *wug* and *dax*) with the prompt may have increased comparisons of *feps* to both *wugs* and *daxes* even when unnecessary. Some participants were redundant, using both rules and similarity to describe the same feature. This occurred more often along the Boolean feature dimension (e.g. "A *fep* is white like a *wug*"), and was less frequent for the continuous feature dimension (e.g. "It has a similar shape to a *dax*; with an indentation on the bottom left and

[2] The main discrepancies were about response relevance to the task, as opposed to the presence of similarity- or rule-like language.

straight lines on the top and bottom right").

These results are not too surprising. The rule based component (i.e., white) is easy to encode and vital to convey the correct concept; whereas, the similarity based component is equally as costly to explain in terms of rules or similarity. Additionally, participants in our task were biased to generalize on the basis of shape. Given this bias, it might be more informative to convey the fill dimension than the shape dimension. Future research will have to tease apart how inductive biases and task demands give rise to verbal descriptions, including how effective participant descriptions convey the information required to identify *feps* and complete our task. In addition, our verbal description task only occurred at the end of the experiment. Future work should examine how verbal descriptions compare and contrast across stages of varying exposure.

### Discussion

The results in Experiment 1 showed that people are able to retain gradience in the representation of concepts where both a Boolean and continuous feature are relevant. Under the assumption that the Boolean feature was evaluated using a rule-based system, and similarly, the continuous feature was evaluated using a similarity-based system, our results provide evidence that rule- and similarity-based systems may be used in conjunction with one another to represent a single concept. In addition, participants demonstrate a change in representation over the course of learning the new concept, and furthermore, this change reflects an initial reliance on similarity, and over time, the introduction of rules into the system. This finding supports previous hypotheses that representations can change with greater exposure and are often initialized using similarity systems.

If in representing concepts, people are combining functions over dimensions as suggested in Experiment 1, it then opens up the question as to how these functional components are combined, in particular in conjunctive relations.

## Experiment 2

Building off of Experiment 1 where we tested how people represented concepts with one Boolean and one continuous feature, Experiment 2 sought to examine how people

represent concepts with two continuous dimensions. By collecting this information of participant responses for this task, we proposed that we would then be able to examine what algorithms participants use to compute conjunctions over each dimension, whether continuous or Boolean.

Under the assumption that the functional components in question are either Boolean (evaluated in a rule-based system) or continuous (evaluated in a similarity-based system), we know the resultant evaluation of the conjunction of these components in some but not all cases. By considering each combination of component type and the resulting value of a conjunction over the components, we may be able to fit a model that can be used to flexibly compute any combination of Boolean or continuous dimensions. The following list examines what we know or do not know of our representation of concepts with two relevant features and a conjunction over any combination of Boolean or continuous components:

(1) 2 Boolean components: From logic, the analysis of conjunctions over Boolean components is trivial; both dimension A and dimension B individually need to be true in order for the conjunction of A and B to be true, otherwise the conjunction would be false.

(2) 1 Boolean component, 1 continuous component: Combining a Boolean component A and a continuous component B would require that A be true in order for a similarity measure of B to be contributed toward the evaluation of the conjunction. Experiment 1 provides evidence that people's conjunctive representations over concepts with a single Boolean and a single continuous dimension are gradient. This finding supports the hypothesis that people preserve the similarity measure in their representation given that the measure over the Boolean component is true.

(3) 2 continuous components: By examining the results of people's representation in such a conjunctive case over two similarity-based measures, we will be able to compare our results with models traditionally used to examine conjunctions over continuous components.

There are a number of models, inspired by the fuzzy logic literature (e.g., Oden, 1977; Budescu, Zwick, Wallsten, & Erev, 1990) that have been proposed to evaluate conjunctions over two continuous components. Here we consider a subset in Table 2.

Importantly each of these models are compatible with results from conjunctions over

| | Model | Function |
|---|---|---|
| (A) | Hard Threshold | $f(a,b) = \begin{cases} 1 & \text{if } a \geq A \ \& \ b \geq B \\ 0 & \text{otherwise} \end{cases}$ |
| (B) | Average | $f(a,b) = \frac{a+b}{2}$ |
| (C) | Euclidean Distance | $f(a,b) = \frac{\sqrt{a^2+b^2}}{\sqrt{2}}$ |
| (D) | Multiplication | $f(a,b) = ab$ |
| (E) | Minimum-Value | $f(a,b) = Min(a,b)$ |
| (F) | Maximum-Vale | $f(a,b) = Max(0, a+b-1)$ |

Table 2

*Proposed models to compute conjunctions over both similarity- and rule-based components For each of the models, we consider the following constraints on the measures:* $0 \leq a \leq 1$, $0 \leq b \leq 1$, *and* $0 \leq f(a,b) \leq 1$.

two Boolean components (1) and all except the Hard Threshold Model (A) are compatible with the qualitative results from conjunctions over one Boolean and one continuous component (2). Importantly, each model makes somewhat different predictions about conjunction over two continuous components (3).

Figure 8 presents the predictions for the six potential models listed in Table 2. We predicted that the Hard Threshold Model (A) would not be able capture the preservation of gradience in mental representations. The other models would be able to capture the gradient mental representations; however, each would predict slightly different quantitative patterns of gradience in the resultant representation. By comparing the results from participant responses in Experiment 2 with the predictions of the models, we are able to analyze which models seem to perform better or worse at capturing people's representations of conjunctions over continuous features.

To test these models, we used a similar artificial concept learning task to that used in Experiment 1. However, rather than using a concept with one Boolean and one continuous feature, we used a concept, also named *fep*, with two different continuous features. We then fitted the models to the results and compared each of the models.

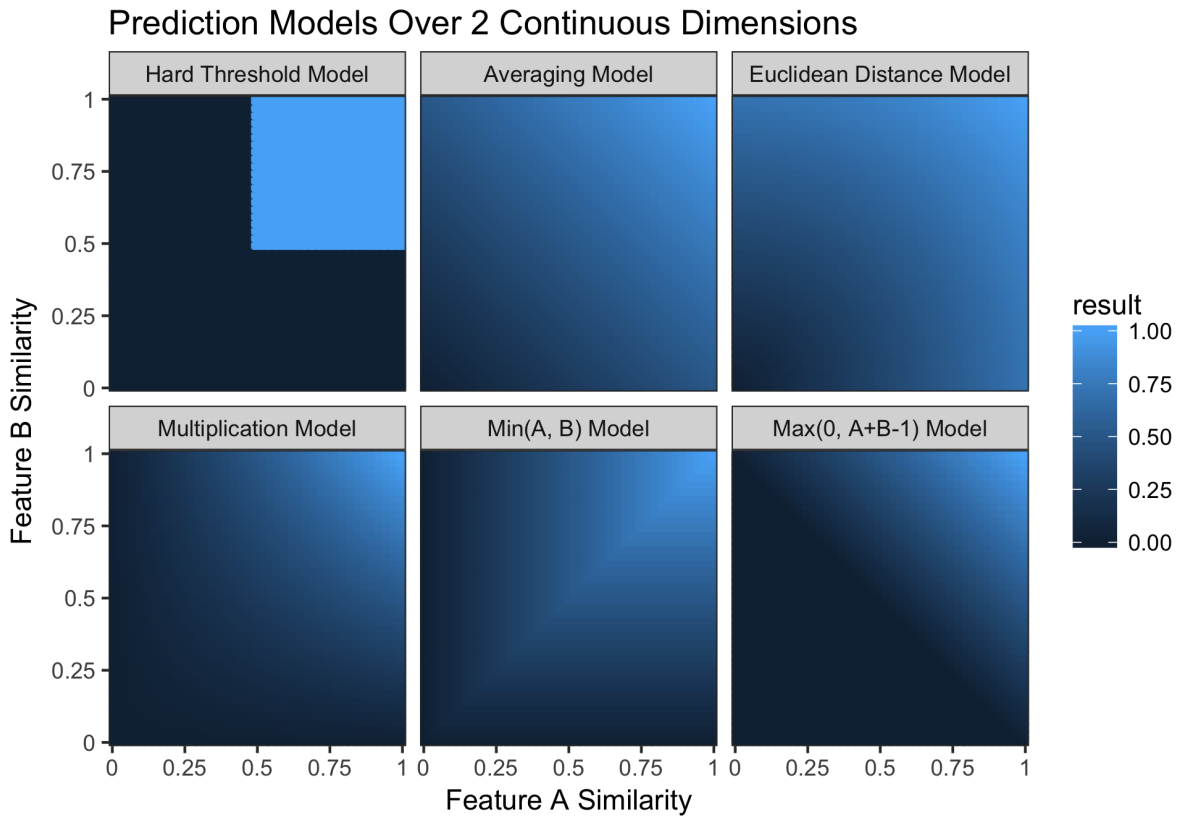


*Figure 8.* Prediction graph for each of the six proposed hypotheses for the representation of concepts with two continuous dimensions. The x- and y-axes represent similarity on two different continuous dimensions. The tile color represents what the model would predict about how likely an object with the corresponding similarity values on each dimension would be generalized to. Lighter blue colors indicate a predicted higher rate of generalization.

## Participants

We recruited 56 participants via Amazon Mechanical Turk. All participants completed all six trials.

## Stimuli

Our stimuli consisted of two relevant continuous dimensions: blob-y shapes and shades of blue. We used ten of the shapes from Experiment 1. Each shape was about equally spaced in measured similarity distance. We additionally generated shades of blue, similarly about equally spaced in similarity distance. We combined each of the ten shapes with each of ten of

the colors, generating a total of 100 unique images.

In order to measure similarity over the shades of blue, we collected norming data ($n = 28$). Two participants were excluded from the analysis because greater than 50% of their responses had similarity rating scores that were less than 5 out of 100. As in Experiment 1, we asked participants to adjust a slider to rate the similarity of each shade of blue to one particular shade of blue. Our pilot studies showed a prominent discretization of the color space when the items tested were too close in color space. Shades of blue with a slight tinge of green were evaluated categorically compared to shades of blue without the tinge of green. We believe that this discretization was caused by a context effect. Because of this, we ran our norming study with 31 stimuli, consisting mostly of blues ranging from blue-green to blue to blue-purple, but also consisting of colors that would be categorized as "green", "purple", and "pink". Similar to Experiment 1, we normalized the ratings by participant, and used the mean value of these normalized similarity ratings to determine the similarity measurement of the stimuli. The results are shown in Figure 9. From these colors, we selected ten among the shades of blue, each about equally spaced in the similarity measurement.

We also collected norming data for the different shapes used in Experiment 2 ($n = 31$), which can be seen in Figure 10. Experiment 2 used a smaller set of shapes (10 shapes) compared to Experiment 1 (50 shapes). This new set of shapes was used in this norming study. The procedure was otherwise the same as in Experiment 1. Two participants were excluded from the analysis because greater than 50% of their responses had similarity rating scores that were less than 5 out of 100.

**Procedure**

The procedure is the same as in Experiment 1, except for the new set of stimuli presented. In Experiment 2, *feps* were considered to be the items that were within the five highest rated similarity instances on both dimensions. In other words, there were a total of 25 items that were considered to be *feps* and 75 items that were not considered to be *feps*. In Figure 8, these are the items highlighted in the Hard Threshold model. The labeled *fep* exemplars were sampled from among these *feps*, and the labeled items presented were

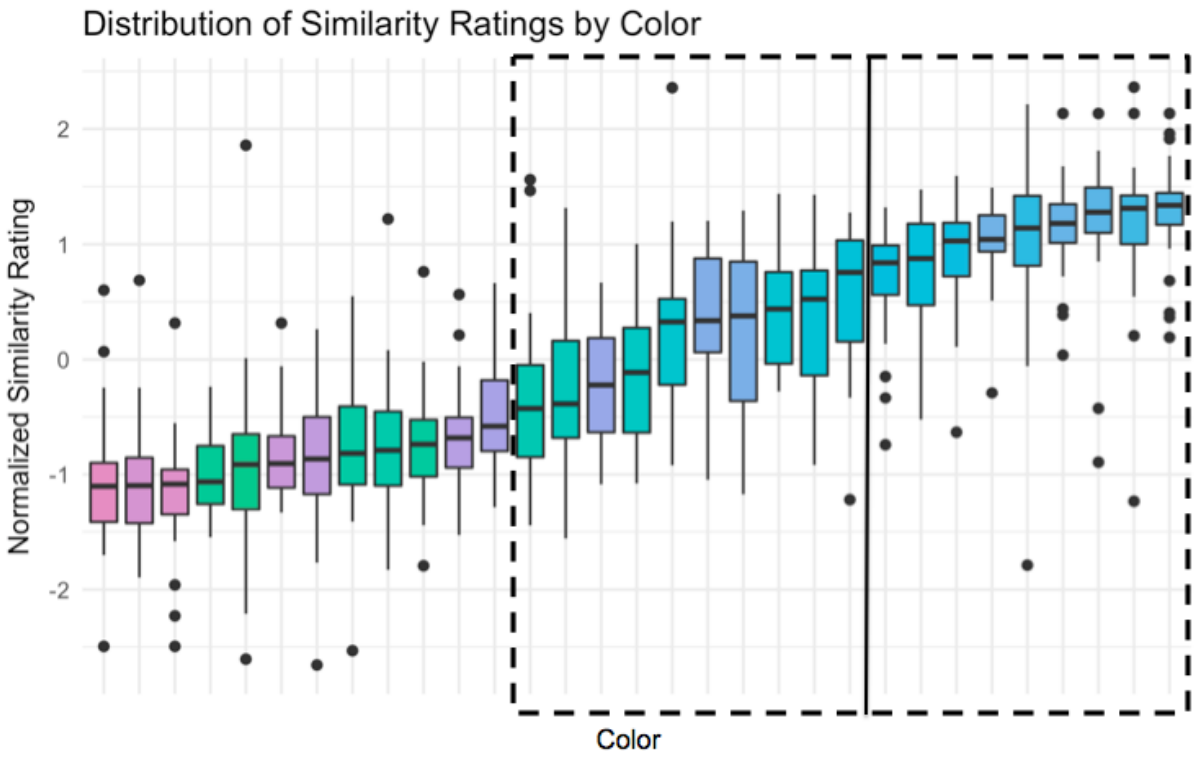


*Figure 9*. Normalized similarity ratings for each of the stimuli colors. The x-axis represents the color items, and the y-axis represents the z-scored similarity ratings by each participant. The color of a given boxplot is the color of the corresponding stimuli. Since additional colors were used in the norming study in order to alleviate context effects, the dashed box indicates the space of stimuli that are used in the concept learning task. The solid vertical line indicates the cutoff in color similarity of items pre-experimentally determined to be *feps*. Slope in mean of similarity ratings indicates that the similarity of the colors is perceived gradiently.

consistent across all participants, though randomized between trials.

As in the description task in Experiment 1, participants were presented with two potential modulus: *wug*, which is most similar on the color dimension but least similar on the shape dimension; and *dax*, which is most similar on the shape dimension but least similar on the color dimension. In Figure 8, the *wug* is represented by the top left corner of each tile plot and the *dax* is represented by the bottom right corner of each tile plot.

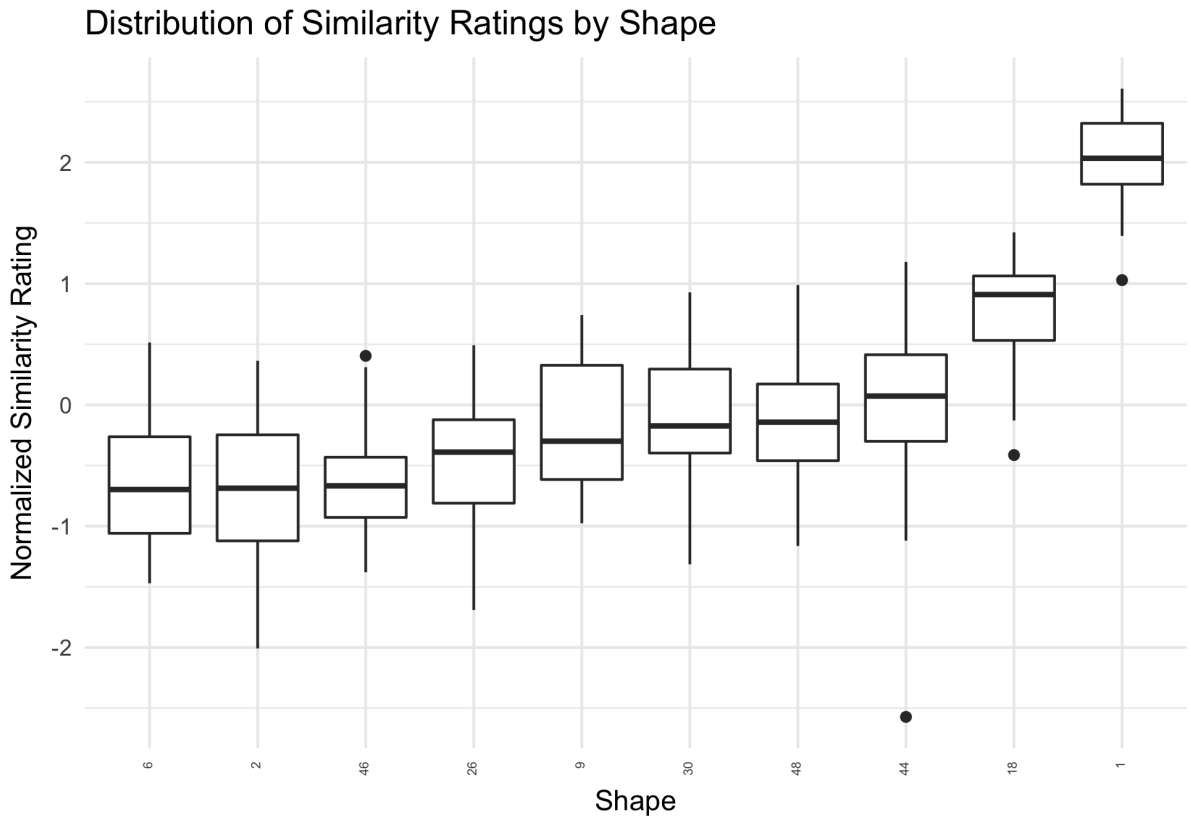


*Figure 10*. Normalized similarity ratings for each of the shapes used for the stimuli in Experiment 2. The x-axis represents the shape items, and the y-axis represents the z-scored similarity ratings by each participant.

**Results**

Figure 11 shows the results from Experiment 2. The x-axis indicates similarity on the continuous shape dimension, and the y-axis indicates similarity on the continuous color dimension. Both dimension are measured using rank, where 1 is the most similar item on a given dimension and 10 is the least similar item on that same dimension. The color of the tile reflects the percent of generalizations to the test items at each of the dimensions. Lighter blue tiles illustrate more generalizations and darker blue tiles, fewer generalizations. Qualitatively, the results show that participants are using both dimensions to determine whether a new instance belongs to the same category. There seems to be a gradience along the color dimension (i.e. top to bottom), but less of a gradience along the shape dimension (i.e. left to right). The shape dimension appears to be almost discrete, such that test items sharing the

exact shape as at least one of the labeled exemplars are often generalized to, whereas test items that do not share the same shape are often not generalized to.

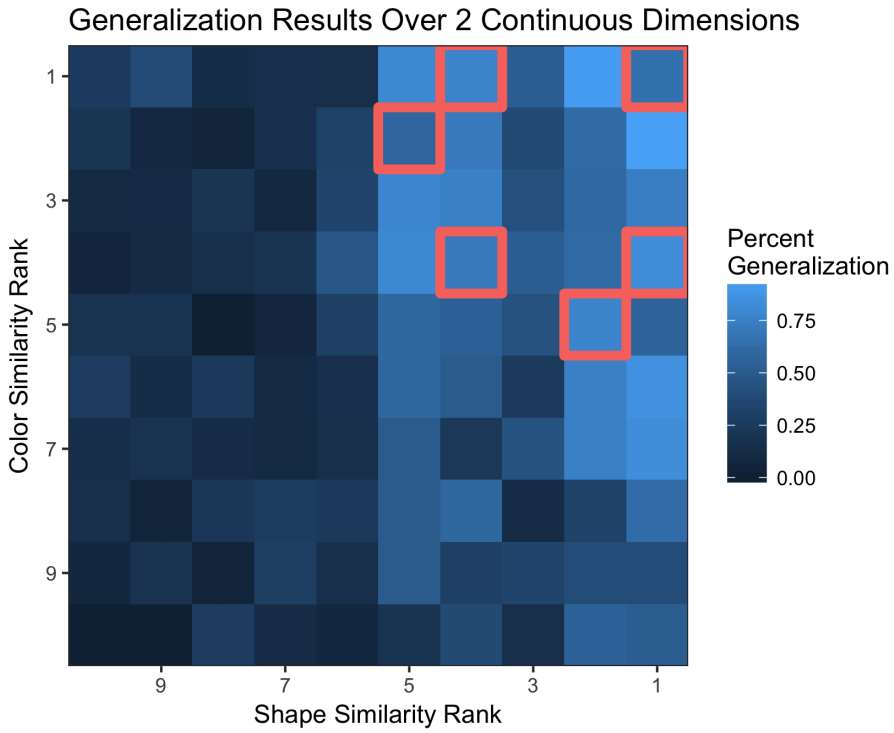


*Figure 11*. Results across participants on the last three trials from Experiment 2 artificial concept learning task, in which objects have two continuous features. The x- and y-axes represent the ranked similarity on two different continuous dimensions. Lighter blue tile colors indicate a higher rate of generalization. The tiles lined in red represent the test items that were also provided as labeled exemplars (i.e. "This is a *fep*").

We analyzed these models by computing the predicted result of the empirical values of the continuous dimensions (i.e. similarity rating for color $C$ and shape $S$), such that $0 \leq C \leq 1$ and $0 \leq S \leq 1$. To convert the similarity values into values between 0 and 1 for each item, the following conversion was used for items across both dimensions:

$$X_i' = 1 - \frac{|X_i - max(X)|}{max(X) - min(X)}$$

The resulting model-predicted value was used as a variable within a logistic mixed-effect regression model, e.g. $Y = \hat{\beta}_0 + \hat{\beta}_1 Min(C_i, S_j)$

We additionally ran regression models on linear combinations of the individual

similarity measures on the dimensions. We examined both the models with and without an interaction. These models are included in Table 3 and are named "Interaction Model" and "Weighted Average Model", respectively. We compared each of the proposed models (Hard Threshold, Averaging, Euclidean Distance, Multiplication, Minimum-Value, Maximum-Value, Weighted Average Model, and Interaction Model), and we used both the Bayesian information criterion (BIC) and log-likelihood to compare the fits of each of the models based on the data. The BIC further allows us to compare models by penalizing for the greater number of parameters within the models, in order to deter overfitting. For the BIC, the lower the value indicates a better fit, and for the log-likelihood, the greater the value indicates a better fit.

| Model | BIC | Log-Likelihood |
|---|---|---|
| Minimum-Value Model | 1842.663 | -906.4784 |
| Euclidean Distance Model | 1843.149 | -906.7216 |
| Weighted Average Model | 1845.051 | -903.9589 |
| Interaction Model | 1850.884 | -903.1625 |
| Multiplication Model | 1851.903 | -911.0985 |
| Averaging Model | 1858.778 | -914.5358 |
| Hard Threshold Model | 1869.275 | -919.7844 |
| Maximum-Value Model | 1874.017 | -922.1553 |

Table 3

*Model comparison using the empirical results from Experiment 2. The BIC and log likelihood were both measured.*

Out of the models tested, the Minimum-Value Model, the Euclidean Distance Model, and the Weighted Average Model have the lowest BIC and highest log-likelihood values, and therefore appear to fit our results the best. In addition, the Hard Threshold and Maximum-Value Models have the highest BIC and lowest log-likelihood, and thus seems to fit our results the least.

Although participants had not been exposed to labeled exemplars with the third most similar color to the prototypical color (i.e. the third row of tiles from the top in Figure 11), the

participants' results indicate that the continuity of gradience is preserved on the color dimension. Alternatively, this did not seem to be the case for the shape dimension, in which the third most similar shape to the prototypical shape (i.e. the third column of tiles from the right in Figure 11) appears to have a lower percentage of generalization compared to its neighboring shapes. The relatively dark column compared to its neighbors provides evidence of this. In other words, there seems to be a caveat in our results, such that participants seem to have a gradient representation of color, but shape has more of a shape-by-shape representation in Experiment 2.

Future studies ought to examine continuous data that can have measurable equidistant steps between items in a given dimension, such as in the case of color in this study. This would more likely guarantee a preservation of gradience in this concept learning task. One potential dimension to examine is the similarity in height to a specific, prototypical height.

**Learning**

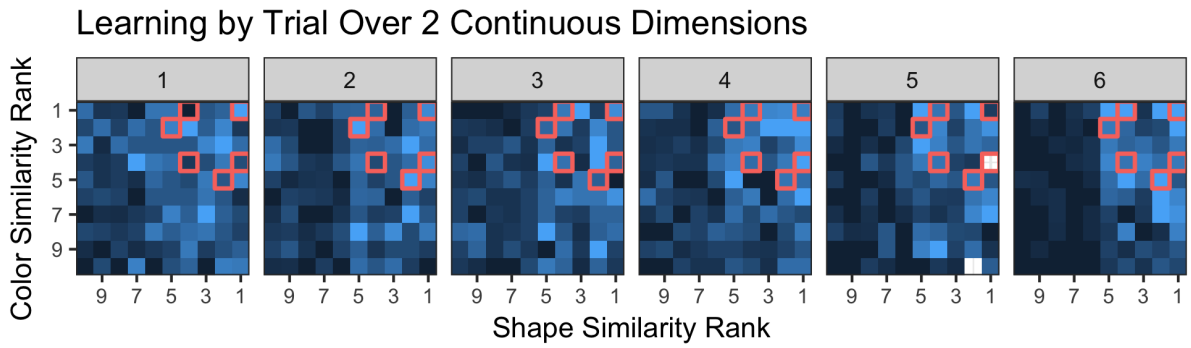


*Figure 12*. Change in representation over the course of six trials in the Experiment 2 artificial concept learning task, in which objects have two continuous features. The x- and y-axes represent the ranked similarity on two different continuous dimensions. Lighter blue tile colors indicate a higher rate of generalization. The tiles lined in red represent the test items that were also provided as labeled exemplars (i.e. "This is a *fep*"). White tiles indicate that no participant saw the given test item during that trial.

As in Experiment 1, the results by trial for all six trials in Experiment 2 indicate a transformation in representation over the course of the experiment. Similar to Figure 11,

Figure 12 looks at the generalization results for test items at each of the different shape and color similarity ranked values. We predicted that participants responses would initially vary across each of the items, and over the trials, would transform toward a more systematic pattern of greater generalization on items more similar on both dimensions (i.e. items closer to the top right corners in Figure 12). Furthermore, we expected the responses to be gradient, such that as items become more similar on each of the dimensions, they are generalized to with higher frequency. Our results show that generalization responses are more random in the earlier trials and become more systematic over the course of the trials. Items more similar on both dimensions do show a pattern of greater generalization. However it is less clear if the generalizations in the later trials show the gradient pattern that we had expected.

**Discussion**

Experiment 2 aimed to test how people represent concepts with two continuous features. Specifically, it sought to examine whether participants evaluated the features compositionally and whether these continuous features maintained a gradient representation when combined in a single concept. The results demonstrate that participants are indeed able to represent concepts with two continuous features compositionally and gradiently; however there are some caveats and proposals of future explorations discussed below.

In testing participants' representations of concepts with two continuous features, Experiment 2 had begun to tackle the question of what computations we use to combine representational components on relevant features. Out of the models tested, we found that the Minimum-Value, Euclidean Distance and Weighted Averaging models best fitted our results from Experiment 2. Meanwhile, the Maximum-Value and Hard-Threshold models performed the worst.

The better fitting models share an intriguing yet intuitive characteristic such that some features are asymmetrically relevant to a given concept's representation. The Minimum-Value model suggests that the least similar feature determines how (un)likely an item belongs to a given category. The Euclidean Distance model suggests that more similar features carry more weight in determining whether an item belongs to a category.

However, Experiment 2's results were unclear as to whether or not gradience had been preserved with the stimuli presented. Evidence of this include what appears to be a categorization of individual items on the shape dimension. For example, the shape third-most (i.e. third column to the right in Figure 11) similar to the given prototype based on the Experiment 1 shape similarity ratings, was generalized to with less frequency compared to the the second- and fourth-most similar shapes (i.e. second and fourth column to the right in Figure 11). The current working hypothesis for the cause of this surprising finding is that none of the labeled exemplars used the third-most similar shape as its shape, unlike the other shapes with close similarity values. However, this issue was not the case for the color dimension, which still demonstrated some degree of gradience. This is in contradiction with the gradience of shape in generalizations seen in Experiment 1 with the same shape stimuli but with fewer unique instances of each shape. One potential explanation for these results is that because fewer shape items were seen in Experiment 2 than in Experiment 1, there may be context effects and the shape bias influencing Experiment 2. Ten shapes may be easier to categorize into their unique categories compared to fifty shapes, as in Experiment 1. In addition, influenced by our prior experience, shapes tend to be a salient feature that facilitates our generation of categories in the world.

Although it is not clear if the results in Experiment 2 demonstrate discretization of the shape dimension, there still seems to be gradience preserved in the evaluation of the color dimension. The results in Experiment 2 illustrate that test items that are true on the discretized "similar-enough" component along the shape dimension and similar on the gradient component along the color dimension are generalized to more frequently, demonstrating the importance of compositionality within the concept's representation. Further research needs to be conducted in order to examine if concepts with two continuous features can be represented compositionally and gradiently across both dimensions. One proposal of such an experiment is replacing the shape dimension with height of a rectangle as one of the continuous features. *Feps* would need to be similar in height to some prototypical height. This new study would allow us to bypass potential issues of the shape bias's effect on discretization of the shape dimension.

## Conclusion

Our results from Experiment 1 have demonstrated that people are able to develop a representation for concepts with both Boolean and continuous features. Similarly, our results from Experiment 2 have testified to the fact that people can represent concepts with two continuous features. People's representations seem to compose rule- and similarity-based representation systems, and people can preserve gradience in compositional representations.

Overall, the results from both Experiment 1 and 2 have provided further evidence in support of hybrid theories of representation. Not only can representation transform from one system to another, but both systems may be used compositionally when necessary to represent a single concept. By examining representation at the feature level, we may examine how two competing systems may not only co-exist but also act to complement one another, as suggested by Heit and Hayes (2011). Future research should explore the factors motivating the use of each system. Where both a purely rule-based and a purely similarity-based system may fail to capture a conjunctive discreteness and gradation of the features, having access to a compositional system such as this points to the flexibility of the human mind's ability to grasp a large variety of concepts.

There still remains an open question about how similarity is measured. Similarity is not simple; for one, it is dependent on context (Tversky & Itamar, 1978). Similarity may be an algorithm that is used after all rule-based algorithms have been exhausted in evaluating a given concept. It may also be the case that similarity can be reduced to a probabilistic function (Goodman et al., 2008). The function likely takes in two parameters: the item in question and either all previously encountered exemplars, a single, prototype, or an abstract representation of the concept. This function may return either a probabilistic value or a Boolean value with some degree of probability. Future research should model how similarity and rule-based systems combine to predict categorization behavior.

In summary, our findings have demonstrated that people flexibly combine rule- and similarity-based representations compositionally within a single concept. Although both systems are often compared and contrasted with one another, our data provided evidence for a larger system that composes similarity and rules as components, depending on the target

concept to be represented. Models of compositionality over components to describe mental representation contributes to a current trend in cognitive science toward the theory of a language of thought (Fodor, 1975).

## Acknowledgments

Thanks to Steve Piantadosi and Frank Mollica for mentoring me throughout this past year. You both have truly helped me to develop into an independent and capable researcher. Thank you to my other committee members, Florian Jaeger and Robbie Jacobs, who have not only provided me with knowledge of statistical analysis and perceptual similarity, but have also given me some of the most memorable tidbits of personal and academic advice throughout my research journey thus far. To Crystal Lee, Chigusa Kurumada, and Linda Liu, you all have been endlessly supportive of me, and I cannot express just how grateful I am for all of your help in getting me through the last couple years. Lastly, thank you to the Computation and Language Lab, the Kinderlab, and the HLP lab for very helpful feedback on experiment design and the logistics of the study, and also, providing me with an environment that I can call my home at the UofR.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth*. Oxford, UK: Wiley.

Budescu, D. V., Zwick, R., Wallsten, T. S., & Erev, I. (1990). Integration of linguistic probabilities. *International Journal of Man-Machine Studies*(33), 657–576.

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, *407*, 630–633.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive science*, *32*(1), 108–154.

Hahn, U., & Chater, N. (1998). Similarity and rules: distinct? exhaustive? empirically distinguishable? *Cognition*, *65*(2), 197–230.

Heit, E., & Hayes, B. K. (2011). Predicting reasoning from memory. *Journal of Experimental Psychology: General*, *140*(1), 76.

Kemler, D. G. (1983). Exploring and reexploring issues of integrality, perceptual sensitivity, and dimensional salience. *Journal of Experimental Child Psychology*, *36*(3), 365–379.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321.

Lassiter, D. (2017). *Graded modality: Qualitative and quantitative perspectives*. Oxford, UK: Oxford University Press.

Margolis, E., & Laurence, S. (1999). *Concepts: core readings*. Cambridge, MA: MIT Press.

Newell, A., & Simon, H. A. (2007). *Computer science as empirical inquiry: Symbols and search*. New York, NY: ACM.

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

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and*

*Cognition*, *14*(4), 700.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological review*, *101*(1), 53.

Oden, G. C. (1977). Integration of fuzzy logic information. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 565–575.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*, 390–398.

Tversky, A., & Itamar, G. (1978). Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 79–98). Hillsdale, NJ: Lawrence Erlbaum Associates.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302.

Werner, H. (1948). *Comparative psychology of mental development.* Oxford, UK: Follett Pub. Co.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7780–7785.