Adults use gradient similarity information in compositional rules

Lauren A. Oey, Francis Mollica, Steven T. Piantadosi
loey@u.rochester.edu, {mollicaf | spiantado}@gmail.com
Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627 USA

Abstract

When learning about the world, we develop mental representations or concepts for things we have never seen. At the same time, we also develop representations for things that are similar to what we have experienced. Traditionally, similarity-based and rule-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*. Using an artificial concept learning task, we test whether people can combine similarity and rules compositionally in order to represent concepts. Our results suggest that people are able to compose similarity and rules when mentally representing a single concept.

Keywords: concept learning, conceptual representation, similarity, rules, generalization

Introduction

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-("like a horse") and rule-based ("striped") 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 a 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 (e.g., Feldman, 2000) typically handle discrete feature values. Rules of this system naturally compose to generate increasingly complex but precise rules. Additionally, rules are easily verbalized making them prime for linguistic transmission. In generating rules, 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.

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.

Additional 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. Goodman, Tenenbaum, Feldman, and Griffiths (2008)'s rational rules model represents a distribution over rules, which can provide gradient beliefs in each possible rule. Heit and Hayes (2011) adopted a model similar to the rational rules model that explicitly incorporates the notion of similarity embodied by the Generalized Context Model (Nosofsky, 1988) as a rule in order to explain inductive reasoning behavior.

In the current literature, it is unclear how the evaluation of gradient features is or could be incorporated in rule-based models. If the evaluation of gradient features can be integrated into a compositional system, will this component cap-

ture patterns of similarity, such as the preservation of continuity within the representation? Our goal in this paper is to see whether people freely combine gradient similarity and discrete rule-like features as components of a larger compositional system. We conducted an experiment to examine whether similarity- and rule-based systems can compose to represent a single concept with both continuous and Boolean dimensions. We argue that both similarity and rules are necessary and can be combined to form mental representations.

Experiment

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 concept, *fep*. Critically the concept being learned contained two relevant features: one that was gradient (i.e. not readily discretized into categories or easily described with language) and one that was binary (i.e. taking discrete categories and easily described with language). Thus, we assumed that the continuous feature would elicit the use of similarity-based representations and the Boolean feature would elicit the use of rule-based representations. 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: they could discretize the continuous feature by, for instance, "coding" the feature as a new binary feature for "similar enough." If participants do this, it would suggest that they are not maintaining the gradient similarity information within the compositional system. A third possible outcome is that participants may not combine systems at all—they may generalize along either the continuous or the discrete dimension, suggesting that participants resist combining systems.

Participants

We recruited 106 participants on Amazon Mechanical Turk. Three participants were excluded because of their failure to complete the task. Two other 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. 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 defining the shapes were determined by a normal distribution and the coordinates of each of the four vertices were determined by independent uniform distributions. That is, 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 itself to forming a prototype, where the shared distribution edge along each vertex acts as this prototype.

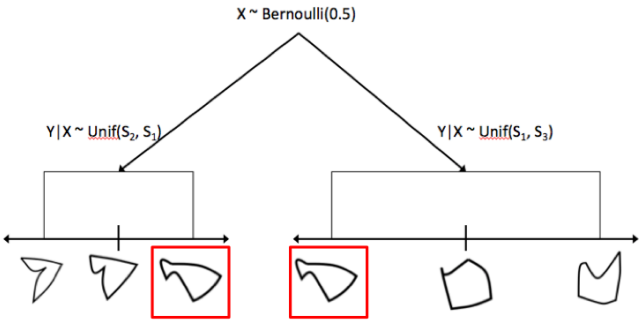

Figure 1: Function used to generate the stimuli shape. Bernoulli distribution followed by a uniform distribution is used to determine what shape is generated. Both uniform distributions share the same shape at the "edge value" (in red box).

Figure 2: Normalized similarity ratings for each of the 50 stimuli shapes. Slope in mean of similarity ratings indicates that the similarity of the shapes is perceived gradiently.

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. Thus, we expect that shape will be perceived as a gradient feature based in similarity to a prototype, and fill color will be perceived discretely as a binary feature commonly studied in Boolean concept learning.

In order to measure how similarly the gradient feature (shape) would be perceived, we first collected norming data ($n = 30$). 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). 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.

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

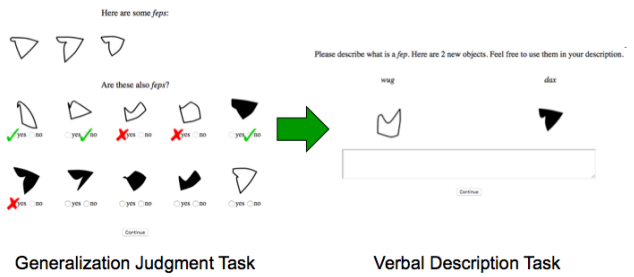

Figure 3: An outline of the experimental setup: (left) the generalization trials where participants respond to which novel shapes are considered *fep*s; (right) the verbal description task with *wug*s and *dax*es labeled for potential use in verbal descriptions.

Procedure

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

Generalization Judgment (Non-Linguistic) Task. Participants were shown exemplars of a novel object *fep*, analogous to how real-world learners might acquire a lexical label (e.g. "kangaroo") by exposure to examples. Participants saw a total of six *fep* exemplars incrementally over six trials. The *fep* exemplars were white and similar to the prototype reference shape used to collect the norming data. The exemplar *feps* were consistent across all participants; although the order in which these exemplars were displayed was randomized for each participant.

On 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. Participants were then shown ten novel test items that were sampled randomly without replacement from the full stimulus set and asked to indicate which were *fep*s by checking yes or no (Figure 3). Participants received feedback after each generalization judgment, being shown either a green check mark if correct, or a red X if incorrect. For the generalization stimuli, *feps* were the 23 white filled shapes that received the highest similarity ratings in the norming study. On average, participants saw 2.8 *feps* per trial (including the exemplar), none of which were the *fep* exemplar itself. Participants made a generalization judgment on each of the 60 test items. Since 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. After labeling the *feps*, participants were shown additional examples.

Verbal Description (Linguistic) Task. At the end of the generalization task, participants were asked to provide a verbal description of a *fep* by typing into a text box. To do so, participants were provided with two novel labelled images to act as potential terms in their descriptions, though they were instructed that they were not required to mention these. One of the objects (*wug*) was white filled but dissimilar in shape to a *fep*. The other (*dax*) was black filled but had a similar shape to a *fep* (see Figure 3).

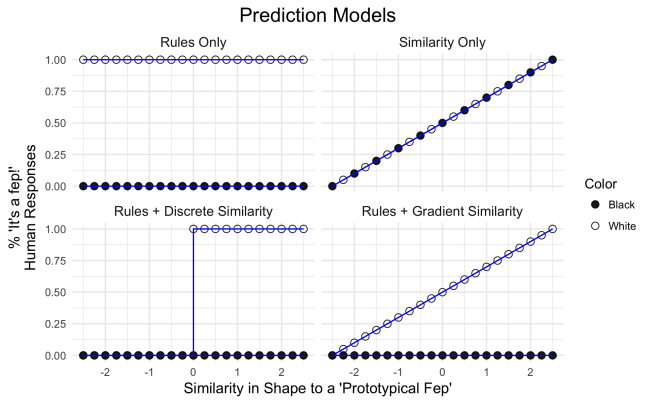

Figure 4: Prediction graph for each of the four proposed hypotheses. The x-axis represents similarity on a scale of -2.5 to 2.5. The y-axis represents the percent of instances that a given item is generalized to (i.e. a test item is predicted to be a *fep*). The line types represent the items with the different discrete features.

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 induced some set of hard and fast rules to capture which items were labelled as *feps* (e.g. flat top and curved bottom), this would suggest that subjects had discretized the continuous space and therefore not combined representational systems. If they did do not articulate these types of rules, this task allows us to use the linguistic data to examine participants' likely representations. For example, participants' descriptions may suggest a similarity-based representation through the use of terms like "like" and "similar to." In general, one can examine the frequency of modal or gradable language (Lassiter, 2017) used in the descriptions in order to distinguish between a similarity function and a probabilistic rule.

Possible Outcomes

Generalization Judgment (Non-Linguistic) Task. We primarily considered four potential outcomes from the non-linguistic experimental task. Figure 4 shows what each possible outcome predicts about how generalization should scale with similarity and the discrete feature. In (a)-(b) 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 but not gradient similarity (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, in (c)-(d) participants may consider both the Boolean and gradient feature when making generalization judgments: **(c)** Boolean and gradient features may 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 differently, depending on the Boolean feature. In other words, similarity would preserve continuity in conjunction with discrete features within a larger compositional structure.

Results

Figure 5 shows the experimental results of the generalization judgments from the last three trials. The x-axis shows the normed similarity ratings for each test item shape. The y-axis shows 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), where subjects show a strong continuous gradient for white items but not for black items. 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 assumed that participants would develop a representation of the concept in the first half of the experiment (i.e. first three trials). We examine the learning aspect separately below.

We also used a logistic mixed-effects regression model to analyze the data from the last three trials. The response variable was whether subjects generalized (yes/no), predicted from the similarity and object fill. Critically, 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 intercepts. The data was analyzed in R using the *glmer* function of the package, *lme4* (Bates, Mächler, Bolker, & Walker, 2014).

Table 1 displays the coefficient estimates and p-values from the model. We found significant effects of fill, similarity, and

	Estimate	z-score	p-value
Intercept	-3.056	-15.466	$p < 0.0001$
Fill	3.065	14.067	$p < 0.0001$
Similarity	0.739	3.059	$p < 0.0030$
Interaction	1.231	3.596	$p < 0.0004$

Table 1: Regression parameters of logistic mixed-effect regression model.

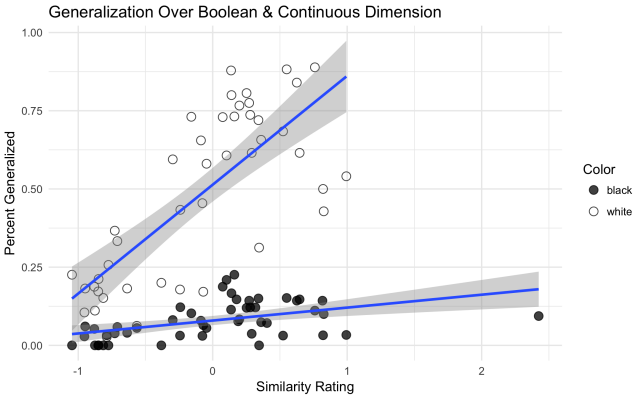

Figure 5: Results from the artificial concept learning experiment across participants over the last three trials of the task. There is an interaction between the Boolean feature (object fill color; represented as point color) and perceived similarity (shape; represented on x-axis) affecting generalization judgments (y-axis).

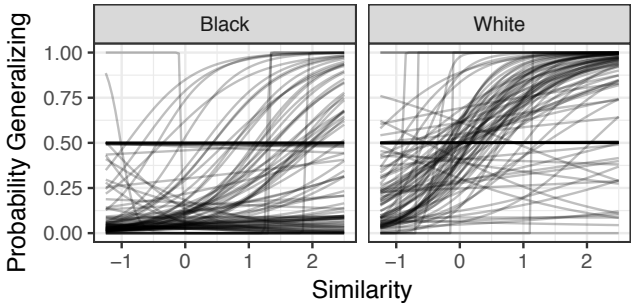

Figure 6: Predictions for individuals generalization probability as a function of similarity for black and white stimuli. The gradient slope for similarity among the white images, suggests participants do not have step-like thresholds for similarity.

the interaction between fill and similarity (Figure 5). These results suggest that people are able to combine similarity- and rule-based systems when developing a representation for the concept *fep* since they not only take into account both features, but how they treat each feature is dependent on the other feature, as reflected by the interaction.

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 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, which shows each individual subjects' classification curve as a function of the similarity. While the judgment is somewhat qualitative, we do not find sharp step-wise functions reflecting discrete similarity thresholds within individuals.

One surprising result that is hinted at in Figure 5, but con-

firmed in our regressions, is that there is a significant effect of similarity within the black filled test items. One potential explanation for this 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 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. The 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 trend shown in Figure 5 begins to emerge. In our data, by around the third trial, participant data is beginning to show a positive slope for the white filled items and a different, flatter slope for the black filled items. Interestingly, the 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) where participants appear to notice the effect of shape before the effect of fill color. Additionally, the slope of the white filled items are shifted upward relative to the black filled items, indicating a smaller tendency to generalize to items that share the same color as well very early. 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. Given our verbal prompts which provided additional labels participants could use, one such predicted verbal response would be "A *fep* is white and like a *dax*." An alternative is that participants could exclusively use similarity-like language (e.g. "A *fep* is like a *dax* in shape and like a *wug* in color.") or use one dimension but not another.

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 2.

As predicted, we find that people do in fact use both rules and similarity-like language (e.g. "A *fep* is more like a *dax* but white."), and this occurred in 36.6% of responses. 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 and discrete; thus, it may be unsurprising that the majority of participants use rule-like language in their descriptions.

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

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

Table 2: Counts of the 101 participant verbal descriptions, categorized by uses of similarity and rule representations in language.

There is one prominent limitation to our verbal description task: Our target concept was not designed to be interrelated (Goldstone, 1996) and so including the moduli (i.e. *wug* and *dax*) with the prompt may have increased comparisons between objects even when unnecessary, as expected by Gentner and Medina (1998). Some participants were redundant, using both rules and similarity to describe the same feature. This occurred more often along the discrete 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 straight lines on the top and bottom right").

These results are not 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 bias 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 occurred at the end of the experiment. Future work should examine how verbal descriptions compare and contrast across stages of varying exposure.

Discussion

Our results have demonstrated that people are able to develop a representation for concepts with both discrete and gradient features. Furthermore, people's representations seem to compose rule- and similarity-based representation systems, reflected both in their generalization patterns and verbal descriptions of these concepts.

Thus, these results 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, perhaps is different domains. Where both a purely rule-based and a

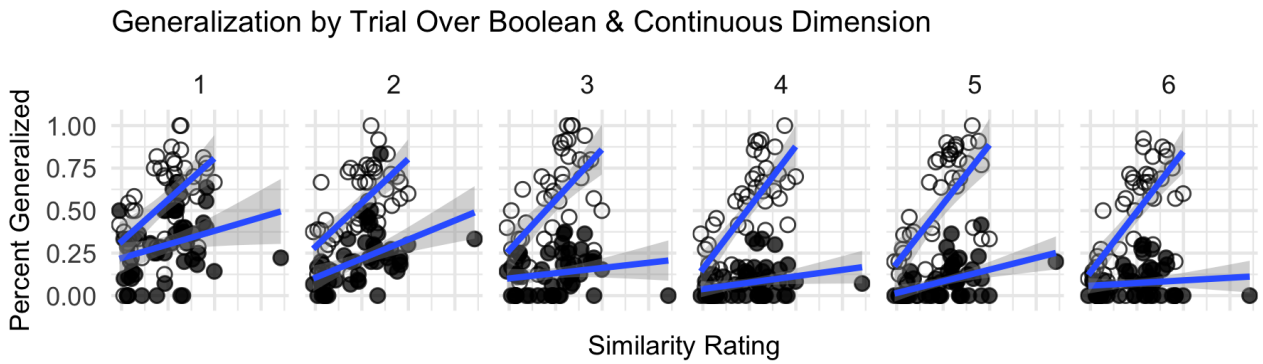

Figure 7: Results of the generalization task broken down by amount of data provided to learners (facets). With increasing data, the interaction slope difference between black- and white-filled items becomes more pronounced.

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.

Note that in spite of our findings, there still remain hard, open questions about how similarity is measured and what exactly it is. Similarity is not simple; for one, it is dependent on context (Tversky & Itamar, 1978). Similarity itself may be a measure 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 characterized as a probabilistic function in a stochastic representation language that is both rule-like and supports gradience through probability calculations (e.g., as in Church: Goodman, Mansinghka, Roy, Bonawitz, & Tenenbaum, 2012), suggesting a natural unifying framework.

Conclusion

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 provides evidence for a unifying 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 Florian Jaeger, Crystal Lee, and the Computation and Language Lab meeting attenders for very helpful feedback on experiment design and the logistics of the study.

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.

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.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2-3), 263–297.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & cognition*, *24*(5), 608–628.

Goodman, N. D., Mansinghka, V., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2012). Church: a language for generative models. *arXiv preprint arXiv:1206.3255*.

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.

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*, *104*(19), 7780–7785.