Title
Exaggeration of Stimulus Attributes in the Representation of Relational Categories

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Du, Yuhui
Hummel, John E.
Petrov, Alexander Alexandrov

Publication Date
2022

Peer reviewed

Exaggeration of Stimulus Attributes in the Representation of Relational Categories

Ava Y. Du (du.618@osu.edu)
Department of Psychology, Ohio State University, 1835 Neil Ave., Columbus, OH 43210

John E. Hummel (jehummel@illinois.edu)
Department of Psychology, University of Illinois, 603 E. Daniel St., Champaign, IL 61820

Alexander A. Petrov (apetrov@alexpetrov.com)
Department of Psychology, Ohio State University, 1835 Neil Ave., Columbus, OH 43210

Abstract

We investigated whether the representation of relational categories is different from that of featural categories. Earlier work has suggested an *extreme-value hypothesis*: when a category is defined in terms of a relation, exemplars with exaggerated values along this stimulus dimension are judged as better members of the category. Featural categories, on the other hand, are not exaggerated. To test this hypothesis, we trained participants to categorize two fictional diseases defined either by a deterministic relation or a deterministic feature. After the categorization task was mastered up to a predefined learning criterion, we provided a graphical user interface that enabled participants to construct good examples of the acquired categories by adjusting the stimulus attributes. We constructed a novel index of relational exaggeration based on residual deviations from a non-exaggerated response strategy. These results supported the extreme-value hypothesis. This replicates and extends an earlier quasi-experimental study (Du et al., 2021).

Keywords: Category learning; relations; typicality; extreme-value hypothesis

Introduction

One can extract information in both featural and relational ways from the same scene. When Ava was chasing her cat while her cat was chasing a squirrel, she had no problem extracting the featural information and categorizing the animals into one group based on their fur coats: *furry(squirrel)* and *furry(cat)*). Once she is above a certain age, Ava would also be able to extract the relational information and categorize the two chasers into one group based on the same role: *chase (Ava, cat)* and *chase (cat, squirrel)*. Nevertheless, the information extracted from the same scene might be represented differently. The features are more likely to be perceivable and constant – the cat's fur coat, for example, produces a characteristic texture pattern on the retina that is visible across many scenes. The relations, on the other hand, require dynamic processing as objects have to be bound with different roles that tend to change rapidly from one scene to the next.

The differences in the representational format for featural and relational information would follow naturally if featural and relational categories were acquired via different mechanisms. It is well established that featural categories can be acquired by associative learning among category labels and a set of featural attributes. This type of learning is captured well by purely connectionist models. On the other hand, relational categories are more likely to be acquired by structure mapping (Gentner, 1983) or schema induction (Doumas et al., 2008; Hummel & Holyoak, 2003). Modeling this kind of learning requires more sophisticated models that combine connectionist and symbol-processing mechanisms (Doumas et al., 2008; Hummel & Holyoak, 2003; Kokinov & Petrov, 2001).

Empirical studies have also found supports for distinct representations of different types of categories. In order to examine whether both common taxonomic and goal-derived categories follow graded structure, Barsalou (1985) asked participants to judge the goodness, central tendency, and ideality of exemplars of these categories. He found that good exemplars of common taxonomic categories are of their central tendencies (e.g., *robin* in *birds*), while good exemplars of goal-derived categories are more of their ideal members that instantiate the goal (e.g., *food with zero calories* in *diet food*). Following Barsalou (1985), distinct representative exemplars in featural and relational categories were observed in many studies (Goldwater et al., 2011; Kittur et al., 2006; Lynch et al., 2000; Rein et al., 2010). Among them, Kittur et al. (2006) trained participants on artificial categories to avoid influence of prior knowledge and explored what would be judged as better exemplars of acquired relational categories. In their task, relational categories were defined by the relative size or darkness between an octagon and a square on each trial. After training, Kittur et al. manipulated whether defining relations and features were consistent with the trained exemplars in a following transfer task (i.e., whether relative size/darkness preserved, and whether the absolute size/darkness were within the range presented during training). Surprisingly, Kittur et al. (2006) found that exemplars with consistent relations and inconsistent features were judged as better exemplars of the categories than exemplars with both consistent relations and features. In other words, extreme relations were found more illustrative of the relational categories even at the sacrifice of familiar features. These studies, taken together, suggested that a good member of at least some relational categories is not close to the prototypes, but it instantiates extreme values of the encoded relations. We refer to this idea as the *extreme-value* hypothesis.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

One limitation of the previous studies was that the dependent measures that they used were poorly suited for testing the extreme-value hypothesis. Rather than measuring the degree of exaggeration of the represented features and relations, most previous studies measured the goodness of a given set of extreme/ideal exemplars. This experimental design does not make clear whether participants represented relational categories by extreme exemplars initially, or just found out that extreme exemplars were good after seeing the given exemplars. To measure directly the degree of representational exaggeration, we created a novel task in which participants were asked to reconstruct good members of given categories (Du et al., 2021).

Our previous experiment (Du et al., 2021) trained participants in artificial categories whose structure was similar to that used in an earlier study by Kittur et al. (2006). Participants learned to categorize two fictional diseases, defined either by a feature or a relation. Each exemplar was depicted by two kinds of artificial cells that varied in four dimensions such as the number of hairs on the surface of the cell or the number of organelles inside it. Each dimension could be acquired as a relation (between two kinds of cells) or a feature (of the more salient kind of cells). As depicted in Figure 1, our previous experiment started with a training phase in which a single deterministic attribute (either a feature or a relation, counterbalanced across the space of eight stimulus attributes) was 100% diagnostic of the correct diagnosis. The other seven attributes were also informative (i.e., 75% diagnostic) of the correct diagnosis, and a perfect accuracy could still be achieved by pooling the information over these partially informative attributes. In such category structure it is possible to achieve perfect accuracy by following multiple distinct strategies. The simplest strategy was to identify the single deterministic attribute and rely on it for classification, ignoring the other attributes. But more complex strategies that "took a majority vote" across several features and/or relations were viable as well. Therefore, our previous experiment included a transfer phase in an attempt to classify the participants according to the type of strategies they adopted (Fig. 1). The experimental session concluded with a novel reconstruction task that measured the degree of exaggeration of good exemplars from the central tendency of training exemplars. As predicted by the extreme-value hypothesis, we found that the encoded relations would tend to be exaggerated in comparison to the encoded features (Du et al., 2021). Nevertheless, we acknowledge the limitation of the quasi-experimental design of our earlier study – the participants sorted themselves into groups depending on the strategies they spontaneously adopted. Because there was no random assignment, it was difficult to draw inferences about causality.

The present study employs a true experimental design that complements and extends the quasi-experimental design of the study of Du et al. (2021). The similarities and differences across the two designs are depicted schematically in Figure 1. In both studies, participants were also asked to categorize two fictional diseases, and one (counterbalanced) attribute was 100% diagnostic of the correct diagnosis. However, no other stimulus attribute has any diagnostic value at all in the present study, whereas many attributes were partially diagnostic in the study of Du et al. (2021). The goal is to constrain the resulting representation, leaving little room for alternative strategies. In the new, fully deterministic category structure the only way for a participant to achieve near-perfect accuracy during the training phase is to identify the task-relevant attribute. This obviates the need for a separate transfer phase. The new experimental design proceeds straight to the reconstruction phase that measures the degree of exaggeration in encoded attributes (Fig. 1).

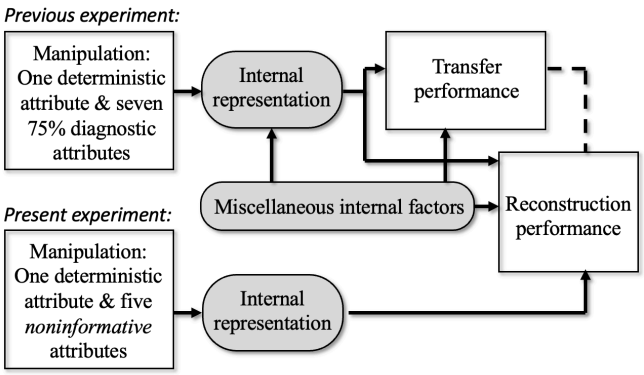

Figure 1: Schematic depiction of our earlier experiment (Du et al., 2021) and the present experiments that complements and extends it. The experimental phases are depicted as white boxes, and hypothesized cognitive structures as shaded ovals. Solid arrows indicate causal links, whereas the dashed line only suggests correlations.

The new design raises an interesting methodological challenge. Because there are two deterministic categories, the overall training distribution of the relevant stimulus attribute is bimodal. Thus, the average value during training is not a good baseline for evaluating the exaggeration of the acquired category representations. Furthermore, the training averages also differed across the two experimental groups (featural versus relational). We propose a novel dependent variable designed to deal with these methodological complications. We defined *a priori* two specific strategies – one featural and one relational – that predict the responses of a hypothetical participant who is performing the task accurately but has zero exaggeration of the task-relevant attribute in their internal representation. The exaggeration of the responses that were produced by the actual participants can thus be estimated by calculating *residuals*. That is, by subtracting the theoretical prediction from the actual observed response.

Overall, we predicted that participants would produce more exaggerated extreme values if the encoded attributes were relations and produce values close to the trained exemplars if the encoded attributes were features.

Method

Participants, Groups, and Inclusion Criteria

There were two main experimental groups - featural and relational - in which the to-be-learned categories were defined by either a feature (F) or a relation (R). Each main group had 3 subgroups counterbalancing the 3 stimulus dimensions. A total of 311 students from the Ohio State University participated in the experiment for course credits. Initially, 244 participants were randomly assigned to the resulting 6 subgroups. Following the predetermined inclusion criteria, we excluded participants who failed more than half of the attention checks, had a median response time less than 0.3 sec, or pressed the same key or same pattern of keys repetitively. To maintain the number of participants in each subgroup comparable after exclusion, we further recruited 67 additional participants. Of all 311 participants, 209 met the inclusion criteria.

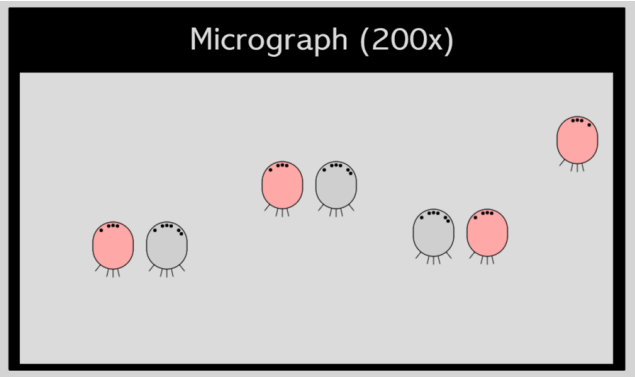

Figure 2: A "micrograph" sample was shown on each trial, containing two kinds of artificial cells that vary in number of cells, number of organelles within each cell, and number of hairs around each cell.

Stimuli and Category Structure

Participants were required to categorize two fictional diseases, Azolitis or Leporidis, based on the features or relations depicted in the "micrograph" on each trial. As Figure 2 shows, each micrograph includes two kinds of artificial cells, grey cells and pink cells, varying in (i) the overall number of cells, (ii) the number of organelles within each cell (shown as dots inside each cell), and (iii) the number of hairs on the surface of each cell (shown as lines). We referred to the grey cells as "healthy" cells and pink cells as "diseased" cells. Depending on the main condition, the diseases were characterized by the absolute features of diseased cells (F) or the relations between diseased and healthy cells (R). Either way, healthy cells were not informative by themselves. Although this rule had not been instructed explicitly, we made the healthy cells less salient in color. Also, the visual display was designed to make counting more efficient in the following ways: (i) the two kinds of cells were presented in pairs whenever possible, (ii) organelles and (iii) hairs were always presented from the center to the sides in groups of three.

As mentioned above, the categories were defined by either the absolute features of diseased cells or the relations between diseased and healthy cells. For example, in the subgroup defined by the absolute number of diseased cells, any micrograph with four diseased cells should be diagnosed as Azolitis, while any micrograph with eight diseased cells should be diagnosed as Leporidis. When the subgroup was defined by the relative number of diseased and healthy cells, on the other hand, any micrograph with fewer diseased cells than healthy cells should be diagnosed as Azolitis, while any micrograph with more diseased cells than healthy cells should be diagnosed as Leporidis. In other words, in each condition, there was either a feature or a relation that was 100% diagnostic of the correct diseases. Moreover, only this defining feature or relation was informative of the diseases. That is, the rest of the features and relations were made to be noninformative, as they would have equal chances of appearing in Azolitis and Leporidis. For instance, Table 1 shows the exemplars presented in the subgroup defined by the relative number of diseased and healthy cells. Accordingly, only the relation in "#Cells" dimension is 100% diagnostic of the categories, while the other relations and features are not diagnostic (i.e., 50% = chance level). Moreover, since those noninformative features and relations were independent, they did not provide extra information even if multiple relations or features were considered together.

Based on the well-defined relations and absolute features of the diseased cells, the absolute features of the healthy cells were then sampled in the following way: considering Weber's law, the slopes of sampling boundaries (as shown in Figure 3) were 1.3 and 1/1.3 so that differences between diseased and healthy cells were easy to notice. At the same time, the sampling range was narrow as well so that differences were not too salient. Figure 3 shows the sampled absolute features of the healthy cells given the relations and absolute features of the diseased cells. Finally, the association between diseases (Azolitis and Leporidis"), relations and features were counterbalanced across participants.

Table 1: Category structure of Azolitis and Leporidis from the subgroup defined by relative number of diseased and healthy cells. ">" and "<" stand for relations between healthy and diseased cells, while "4" and "8" stand for absolute features of diseased cells.

Dimensions	Exemplars of Azolitis							
#Cells	>,4	>,4	>,4	>,4	>,8	>,8	>,8	>,8
#Organ.	>,4	>,8	<,4	<,8	>,8	>,4	<,8	<,4
#Hairs	<,8	<,4	>,8	>,4	>,8	>,4	<,8	<,4
	Exemplars of Leporidis							
#Cells	<,4	<,4	<,4	<,4	<,8	<,8	<,8	<,8
#Organ.	>,8	>,4	<,8	<,4	>,4	>,8	<,4	<,8
#Hairs	>,4	>,8	<,4	<,8	<,4	<,8	>,4	>,8

Procedure and Scoring

Participants were instructed that there were two fictional diseases, Azolitis and Leporidis, and their task on each trial was to make a diagnosis based on the micrograph taken from an organ of a given patient with either Azolitis or Leporidis. Then, a micrograph was presented, and the participants were instructed that grey cells refer to healthy cells and the two diseases were characterized by the presence of diseased cells stained in pink. Moreover, the three stimulus dimensions that might be relevant to the diagnostic criteria were pointed out explicitly, while no hint was given on the defining feature or relation, or the general usefulness of relations vs. features.

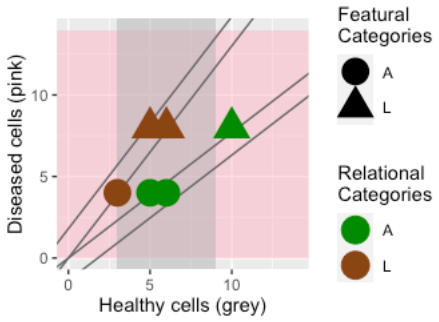

Figure 3: Illustration of sampling absolute features of healthy cells given pre-defined relations and absolute features of the diseased cells. Light pink and grey shades indicate the available ranges of diseased and heathy cells during reconstruction.

After instruction, participants started the training phase, in which they were presented with a sequence of micrographs and asked to make a two-alternative forced choice by pressing A (for Azolitis) or L (for Leporidis) based on the micrograph on each trial. Feedback on whether the diagnosis is correct or not was given right after each choice. Since no hints were given on the defining feature or relation, participants had to make random guesses at first, but they learned from the feedback and gradually mastered the diagnostic criteria for each disease. Trials were organized in blocks of 16 exemplars (as listed in Table 1) in random order. Starting from the second block, participants were allowed to enter the reconstruction phase as long as they achieved the learning criterion, which required correctly diagnosing at least 15 out of 16 micrographs in a block for two consecutive blocks. Otherwise, participants would run through all 320 trials to enter the reconstruction phase.

After the training phase, all participants entered the reconstruction phase. On each trial, they were asked to construct a micrograph as a good example of either Azolitis or Leporidis. There were 5 reconstruction trials for Azolitis and 5 trials for Leporidis. The first two trials included one for Azolitis and one for Leporidis, in a random order, and the rest eight trials were also presented in a random order. Each trial started with healthy cells presented in the micrograph only. Participants adjusted the absolute features of diseased cells using a graphical user interface with three sliders. The absolute features of healthy cells were sampled from 3 to 9 (as shown in the first column in Table 2). The sliders defined a range from 0 to 14 for possible features of diseased cells[1], as grey and pink shades shown in Figure 3. Before completing each trial, participants were forced to either move the sliders to produce the features or relations or click an "Ignore" button drawn next to each slider to indicate that the particular dimension was not included in their diagnosis criterion.

Table 2: Illustration of ideal values of diseased (pink) cells, as zero degrees of exaggeration, given different presented values of healthy (grey) cells and attribute encoded (i.e., Relation or Feature). Values with asterisks in indicate that the values that were presented during training. Association between category labels and relative or absolute magnitudes were counterbalanced to be consistent with training.

	Azolitis		Leporidis	
#Grey	#Pink-Feature	#Pink-Relation	#Pink-Feature	#Pink-Relation
3*	4*	2	8	4*
4	4	3	8	5/6
5*	4*	4*	8*	8*
6*	4*	4*	8*	8*
7	4	5	8	9
8	4	6	8	10
9	4	7	8	11

In order to examine the degree of exaggeration, we compared the reconstructed values of diseased cells to the "ideal" values of diseased cells. Here, the "ideal" values were designed to have zero exaggeration relative to the values that were encountered during training. If there is no difference between reconstructed and ideal values, then we would conclude that there is zero degrees of exaggeration in the encoded representation. However, given that the trained features (i.e., absolute features of diseased cells) and the trained relations (i.e., relative differences between diseased and healthy cells) were different measures, we defined different ideal values for defining features and relations. Table 2 lists all possible values used to depict healthy cells during reconstruction, together with corresponding ideal values for encoded features and relations. The ideal values for encoded features were straightforward, as four and eight were the only values that participants had been trained on for each disease, they would serve as the ideal values for defining features. The ideal values for encoded relations, on the other hand, depend on the given healthy cells on each reconstruction trial as well as the trained differences between diseased and healthy cells. Here,

[1] Except for the overall number of cells dimension, where the range was from 1 to 14.

we quantified the differences by Michelson contrast[2]. Based on the given healthy cells and trained differences (from -.200 to -.111 and from +.091 to +.231), the expected ideal values of diseased cells for encoded relations were calculated (i.e., "#Pink-Relation columns) so that the ideal differences were within the range of trained differences.

Based on this a priori definition of lack of exaggeration, we defined the measure of our main interest: *sign adjusted residuals*, as the differences between reconstructed and ideal values. The sign of the residuals was adjusted so that positive residuals always indicate exaggeration in the direction that agrees with the encoded relations. Finally, we averaged the residuals across five reconstruction trials and two diseases to produce a single dependent measurement per participant.

Results and Discussion

Training Phase

Seventy-one participants reached the learning criterion in the relational condition and 79 did so in the featural condition. In comparison with the previous study (Du et al., 2021) in which only half of the participants passed the learning criterion, more than two-thirds of participants passed it here. This suggests that a simplified environment with fewer informative distractors is more learnable.

Following Kittur et al. (2006), we treated the participants who failed to achieve the learning criterion as if they had passed it on the last trial. With those participants included, no statistically significant difference was found in terms of the trials used to reach the learning criterion (t (204.42) = -.40, p = .69) between participants in the relational condition (M = 160.78, SD = 118.56) and those in the featural condition (M = 167.09, SD = 109.03). This result is consistent with that for deterministic conditions in Kittur et al. (2004), suggesting that the features and relations are comparably learnable as long as both of them are deterministic.

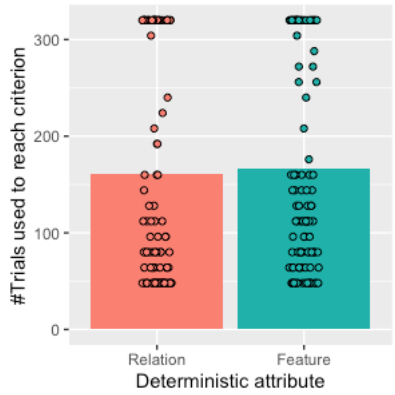

Figure 4: Average and individual number of trials used to reach the learning criterion in relational and featural conditions

Reconstruction Phase

The following analysis include both participants who reached the learning criterion and those who failed the learning criterion.

The extreme-value hypothesis predicts that the learned relations would be exaggerated. In our study this translates into sign adjusted residuals that are significantly greater than zero. Meanwhile, we also predicted that the learned features would be close to the values observed during training. This translates into sign adjusted residuals that are not significantly different from zero. When the learning criterion was not reached, however, we predicted that the reconstructed values would be randomly selected within the range of 0 to 14. Under our scheme for calculating residuals, the sign adjusted residuals of participants who respond at random were predicted to be significantly smaller than zero.

The results were consistent with the extreme-value hypothesis. The average sign adjusted residuals of encoded relations (M = .61, SD = 1.62) was significantly larger than zero (t (70) = 3.20, p = .001). This suggests that when a category representation is organized in terms of a relation, this relation tends to be exaggerated when the participant is asked to produce a new exemplar of the category. On the other hand, the average sign adjusted residuals of encoded features (M = .02, SD = 1.50) were not significantly different from zero (t (78) = .13, p = .45). This suggests that when a category representation is organized in terms of a feature and the participant is asked to produce a new exemplar, this feature will be set to the value that was typical during training. In addition, the average sign adjusted residuals of the participants who failed to reach the learning criterion (M = -2.35, SD = 1.48 in Relation; M = -1.76, SD = 1.78 for Feature) were both significantly smaller than zero (t (31) = -8.99, p < .001 for Relation; t (26) = -5.15, p < .001 for Feature).

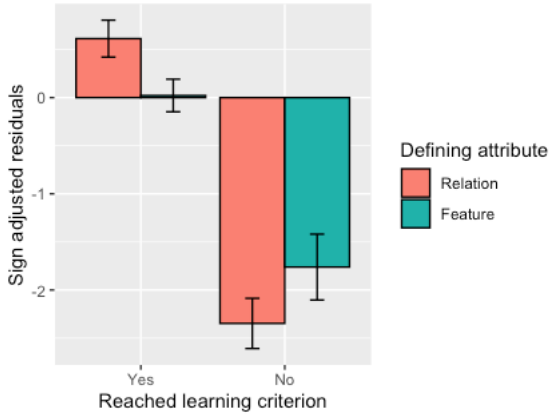

Figure 5: Average sign adjusted residuals of encoded relations versus average sign adjusted residuals of encoded features. Error bars indicate the standard errors.

[2] Michelson contrast is defined as (#pink - #grey) / (#pink + #grey).

General Discussion

We extended and complemented the quasi-experimental study of Du et al. (2021). The current study has a true experimental design that supports inferences about causality. The results of the new experiment replicate and reinforce our earlier results. Both correlational (Du et al., 2021) and causal evidence (present study) support the extreme-value hypothesis. Concretely, participants who spontaneously chose a relational categorization strategy (Du et al., 2021) or were experimentally induced to adopt one (present study) tended to exaggerate the task-relevant stimulus dimension. This exaggeration occurred only in the relational condition – whether it was adopted spontaneously or because of random assignment of participants into experimental groups. There was no evidence of analogous exaggeration of other stimulus attributes that were not relevant for the categorization task. Importantly, there was no evidence of exaggeration of any stimulus attribute in the featural condition. The selective nature of these exaggeration results makes them hard to account for in terms of alternative explanation involving opponent categories (e.g., Davis & Love, 2010).

The evidence obtained using our novel reconstruction task adds to the literature suggesting that the representation of at least some relational categories is different from that of featural categories (Goldwater et al., 2011; Kittur et al., 2006; Rein et al., 2010). That is, while featural categories are represented as the central tendencies of the observed exemplars, relational categories are better represented as the extreme values that instantiate the relations.

As a topic for future research, the reconstruction task can also be used to test the relational learning processes. For example, it is claimed that relational concepts can only be learned when they are deterministic (e.g., Kittur et al., 2004). However, relational learning can be achieved when a higher or lower level of deterministic relation is available (Jung & Hummel, 2015a, 2015b). A reconstruction task can help examine the representation adjustment through the learning process, like whether relations are treated as features somehow to acquire the probabilistic structure.

References

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629-654.

Davis, T., & Love, B.C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science, 21*, 234-242.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, 115, 1–43.

Du, A. Y., Hummel, J. E., & Petrov A. A. (2021). Probing the Mental Representation of Relation-Defined Categories. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar-Raible (Eds.) (2021). Proceedings of the 43rd Annual Meeting of the Cognitive Science Society (pp. 2024-2030). Vienna, Austria.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive science*, *7*(2), 155-170.

Gentner, D. & Kurtz, K. J. (2005). Relational categories. In W. K. Ahn, R. I. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.) *Categorization inside and outside the laboratory*. Washington, DC: APA.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–28.

Goldstone, R.L., Steyvers, M., & Rofosky, B.J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition, 31*, 169– 180.

Goldwater, M. B., Don, H. J., Krusche, M. J., & Livesey, E. J. (2018). Relational discovery in category learning. *Journal of Experimental Psychology: General, 147*, 1-35.

Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition, 118*(3), 359-376.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220–264.

Jung, W. & Hummel, J. E. (2015a). Making probabilistic relational categories learnable. *Cognitive Science*, *39*(6), 1259-1291.

Jung, W. & Hummel, J. E. (2015b). Revisiting Wittgenstein's puzzle: Hierarchical encoding and comparison facilitate learning of probabilistic relational categories. *Frontiers in Psychology*, *6*:110, 1-11.

Kim, S. W. & Murphy, G. L. (2011). Ideals and category typicality. *Journal of Experimental Psychology:Learning, Memory, and Cognition, 37*(5), 1092-1112.

Kittur, A., Holyoak, K. J., & Hummel, J. E. (2006). Ideals aren't always typical: Dissociating goodness-of-exemplar from typicality judgments. In R. Sun & N. Miyake (Eds.), *Proc. Of the 28th Annual Conference of the Cognitive Science Society* (pp. 429–434). Mahwah, NJ: Erlbaum.

Kittur, A., Hummel, J. E., & Holyoak, K. J. (2004). Feature- vs. relation-defined categories: Probab(alistic)ly not the same. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proc. Of the 26th Annual Conference of the Cognitive Science Society* (pp. 696–701). Mahwah, NJ: Erlbaum.

Kokinov, B., & Petrov, A. (2001). Integrating memory and reasoning in analogy-making: The AMBR model. *The analogical mind: Perspectives from cognitive science*, 59-124.

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. Memory & Cognition, 28, 41-50.

Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psych. Review*, *92*, 289-316.

Rein, J. R., Goldwater, M. B., & Markman, A. B. (2010). What is typical about the typicality effect in category-based induction? *Memory & Cognition, 38*(3), 377-388.