

Topological Relations Between Objects Are Categorically Coded



Andrew Lovett¹ and Steven L. Franconeri²

¹U.S. Naval Research Laboratory, Washington, DC, and ²Department of Psychology, Northwestern University

Psychological Science
1–11
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797617709814
www.psychologicalscience.org/PS
SAGE

Abstract

How do individuals compare images—for example, two graphs or diagrams—to identify differences between them? We argue that categorical relations between objects play a critical role. These relations divide continuous space into discrete categories, such as “above” and “below,” or “containing” and “overlapping,” which are remembered and compared more easily than precise metric values. These relations should lead to categorical perception, such that viewers find it easier to notice a change that crosses a category boundary (one object is now above, rather than below, another, or now contains, rather than overlaps with, another) than a change of equal magnitude that does not cross a boundary. We tested the influence of a set of *topological* categorical relations from the cognitive-modeling literature. In a visual same/different comparison task, viewers more accurately noticed changes that crossed relational category boundaries, compared with changes that did not cross these boundaries. The results highlight the potential of systematic exploration of the boundaries of between-object relational categories.

Keywords

categorical perception, visual comparison, spatial relations, topological relations, sequential same/different task, open data

Received 4/7/16; Revision accepted 4/23/17

When people interpret or create visual explanations—for example, graphs, diagrams of physical systems, or depictions of biological processes—visual comparison plays a central role. In each of these cases, people explore, understand, and explain by identifying commonalities and differences between the structure of what they are currently seeing and what they have seen before. In a graph, for example, you can easily pick out an interaction effect present in the data (see Fig. 1, top row). In the depiction of a bacterium entering a cell, you can see the process unfold as a series of discrete steps (see Fig. 1, bottom row). A better understanding of how people make these types of comparisons—what, and how much, they can store and compare—would inform understanding not only of how the human visual system encodes and compares visual structures, but also of how to design visual depictions that facilitate this process for students and scientists.

In this article, we report four experiments demonstrating that people code relational structure *categorically*. Additionally, the experiments begin to catalogue

the boundaries of these categories. Categorical coding discards metric precision in favor of more efficient discrete representations of critically diagnostic features (Biederman, 1987). Categorical coding is typically demonstrated by showing that changes to a stimulus are easier to detect when they cross a category boundary than when they do not, even when the amount of metric change is equivalent. For example, the interaction depicted in the top row of Figure 1 is easier to detect in the left graph, compared with the right graph, because in the left graph, there is a reversal of the relative heights of the two bars, a categorical change. Similarly, we would predict that in the bottom row of Figure 1, the first two steps in the bacterium’s path are more distinguishable in the left diagram, compared with the right diagram, because in the left diagram, the bacterium

Corresponding Author:

Andrew Lovett, 501 Holland Lane, Apt. 813, Alexandria, VA 22314
E-mail: andrew@cs.northwestern.edu

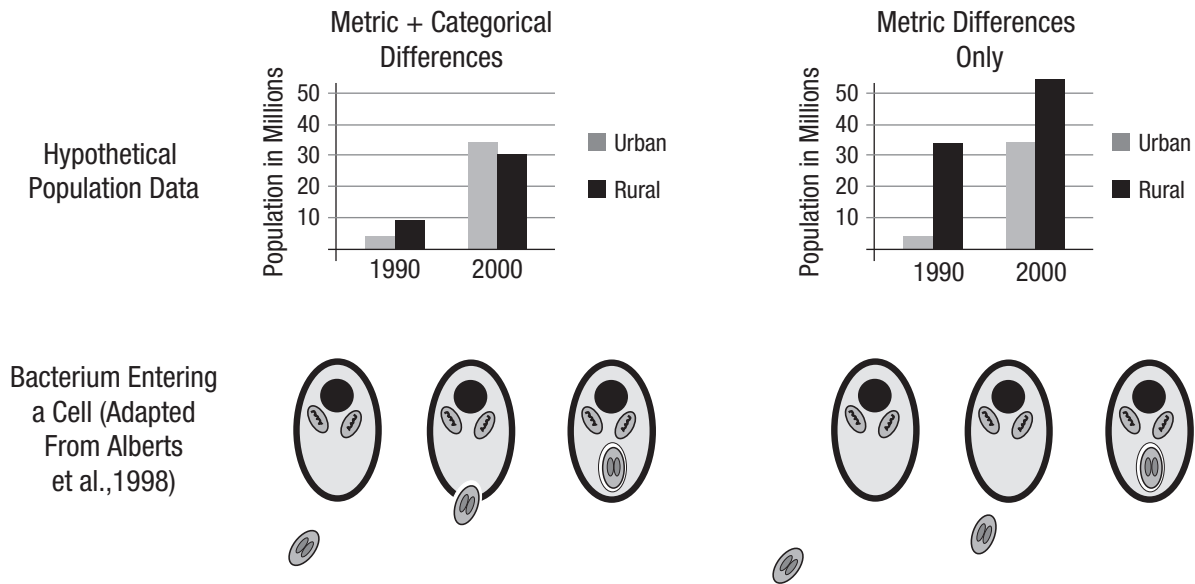


Fig. 1. Examples of visual comparisons with both metric and categorical differences (left column) and with metric differences only (right column). The graphs in the top row depict interaction effects in hypothetical population data; although the metric differences depicted are the same in the two graphs, the interaction is easier to detect in the graph on the left because there is a reversal of the relative heights of the two bars, a categorical change. The diagrams in the bottom row depict a bacterium entering a cell. Although the metric differences between the first and second steps in the bacterium's path are the same in the two diagrams, those steps are more distinguishable in the diagram on the left because the bacterium has crossed the perimeter of the cell, a categorical change in the relationship between the two objects.

has crossed the perimeter of the cell, a categorical change in the relationships between the two objects.

Categorical coding has been well documented in both perception and cognition. For example, it is easier to distinguish blue from green than to distinguish two shades of green, even with metric distance in color space controlled (Bornstein & Korda, 1984). There are similar effects for facial expressions (Etcoff & Magee, 1992), auditory phonemes (MacKain, Best, & Strange, 1981), geometric shapes (Amir, Biederman, Herald, Shah, & Mintz, 2014), and size categories (Kosslyn, Murphy, Bemserfer, & Feinstein, 1977). In addition, it is easier to distinguish two objects when there is a categorical difference in the relationship between the objects' parts than when there is no categorical difference (Hummel & Stankiewicz, 1996; Rosielle & Cooper, 2001).

There are far fewer studies exploring categorical coding of *between-object* relations. When viewers are asked to remember the position of a dot within a circle, their memories are biased by the quadrant of the circle in which the dot was located, which suggests that the circle's quadrants are coded categorically (Huttenlocher, Hedges, & Duncan, 1991). Similarly, when viewers compare two images each containing a dot and a cross, they are better at detecting changes to the dot's position that place it in a new quadrant of the cross, as opposed to an equivalent metric change that keeps the dot in

the same quadrant of the cross (Kranjec, Lupyan, & Chatterjee, 2014). Finally, when participants are asked to choose which of two images better matches a sample image, they select the correct match more quickly when the distractor image differs from the sample along a category boundary of touching versus not touching (Kim & Biederman, 2012). Although these studies are promising, they have only begun to explore the full range of categorical between-object relations used in visual comparison.

Our goal in the experiments reported here was to systematically test a broad suite of potential relationships, by taking inspiration from computational models of human vision and spatial reasoning. Specifically, we turned to the literatures on how people compare categorical relations (Doumas & Hummel, 2013; Falkenhainer, Forbus, & Gentner, 1989; Hummel & Holyoak, 1997; Gentner, 1983; Larkey & Love, 2003) and how they perform on visual problem-solving tasks (Carpenter, Just, & Shell, 1990; Cirillo & Ström, 2010; Kunda, McGregor, & Goel, 2013).

One set of models (Lovett & Forbus, 2011a, 2017; Lovett, Tomai, Forbus, & Usher, 2009) builds on both literatures, using categorical relations to compare images and solve problems. Inspired by behavioral studies (e.g., Huttenlocher et al., 1991) and previous computational models, these models posit a vocabulary of categorical relations that could be used to solve a

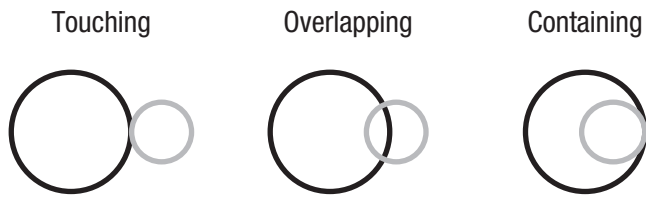


Fig. 2. Proposed topological relations for two objects: touching, overlapping, and containing. Note that whenever two objects are overlapping, they must also be touching, and when one object is containing another, they may also be touching.

diverse set of visual comparison problems. For example, the models build on computational spatial reasoning (Klippel, Yang, Wallgrün, Dylla, & Li., 2012; Randell, Cui, & Cohn, 1992) to propose three topological relations (Fig. 2): *touching* (i.e., edges that touch each other), *overlapping* (i.e., a common region within two

objects), and *containing* (i.e., one object being contained within another). The models' patterns of performance mirror those of humans across multiple tasks. This suggests that the relations and category boundaries in the models may be similar to those used by humans.

In the experiments reported here, we set out to test whether humans are sensitive to the three proposed topological relations. We view these experiments as a first step toward a systematic exploration of the modeled vocabulary of categorical relations. We tested whether topological relations are categorically coded by measuring viewers' sensitivity to changes that either crossed, or did not cross, the hypothesized categorical boundaries. For this purpose, we designed sequences in which a small circle's position relative to a larger circle shifted horizontally by the same distance from each frame to the next (Fig. 3a). In some cases, the positional change resulted in a categorical change in

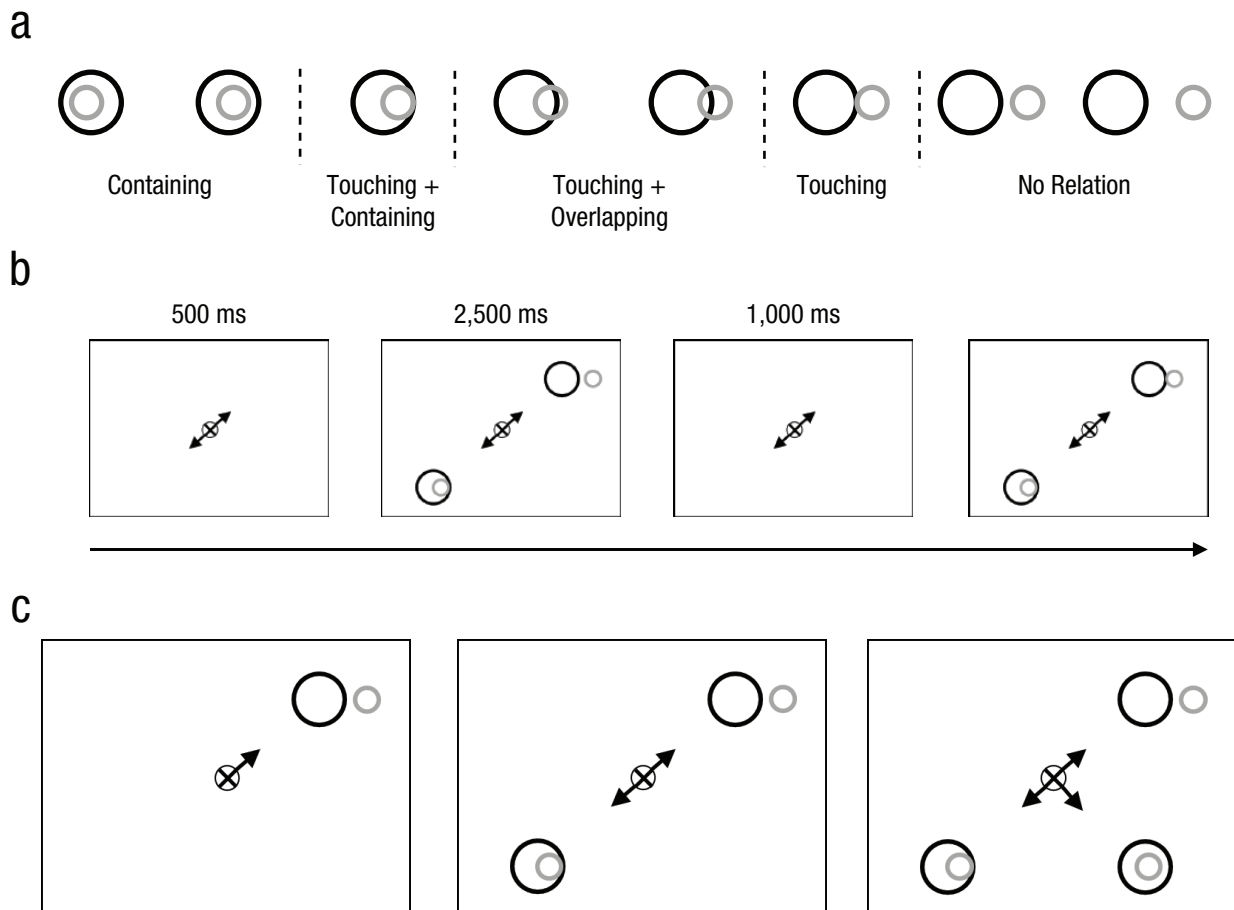


Fig. 3. Illustration of the experimental method. The diagrams in (a) show the eight circle pairs and intervals between adjacent pairs that were used in the experiments. The larger circle could contain the smaller circle, both contain and touch the smaller circle, touch and overlap the smaller circle, touch the smaller circle, or have no topological relation with the smaller circle. The dashed vertical lines indicate the intervals that were hypothesized to introduce a categorical, rather than merely metric, change in the topological relations between the two circles. The trial sequence in (b) illustrates a typical trial with two circle pairs. Each trial began with an on-screen arrow or arrows indicating where participants should attend. After 500 ms, a circle pair appeared in each cued quadrant; 2,500 ms later, the pairs disappeared. After another 1,000 ms, the pairs reappeared, and the task was to report whether the first and second displays were the same or different. The diagrams in (c) illustrate the manipulation of set size (i.e., displays with one, two, or three circle pairs).

the relations between the circles; for example, when the change caused the two circles to touch, that created a “touching” relation. We predicted that participants would notice a change more readily when it crossed a category boundary than when it did not.

In Experiments 1a and 1b, we varied the number of circle pairs in the display (Fig. 3c) to measure whether the extent to which participants rely on categorical (vs. metric) information varies according to how many elements they must keep track of. In Experiment 2, we varied whether the circles were filled or empty, as a first step toward exploring the visual features that might guide the encoding of topological relations.

General Method

Materials

All four experiments used the eight circle pairs and seven intervals between adjacent pairs shown in Figure 3a. Four of the intervals introduced a categorical change (e.g., a change from containing to touching plus containing), and three were purely metric (e.g., the circles in both of the adjacent pairs were both touching and overlapping). To equate the numbers of categorical and metric intervals, we included eight intervals in the experimental designs, randomly choosing a repetition of either the touching-plus-overlapping or the no-relation interval as the eighth interval. Each trial used either the circle pairs in Figure 3a or their mirror reflections (i.e., in which the small circle was to the left of the large circle). To decrease the chance that the two circles in a pair would be perceptually grouped as a single object, we assigned the large and small circles different colors, either red and green or blue and yellow, and the smaller circle was always drawn in front of the larger circle. The set of colors used and assignment of the colors to the large and small circles were counterbalanced across participants.

Procedure

The experiments used a sequential same/different paradigm (Fig. 3b). On-screen arrows cued participants to attend to one, two, or three quadrants. After 500 ms, a circle pair appeared in each cued quadrant and remained visible for 2,500 ms. This display was followed by a 1,000-ms delay, consisting of a 250-ms mask (in which each large circle was covered by an assortment of randomly placed small circles) and a 750-ms blank screen (with just the arrows), before the circle pair or pairs reappeared. Participants’ task was to report whether the first display differed from the second. They pressed one key on the computer keyboard if the pair

or pairs of circles were the same, and another key (with their other hand) if they noticed any difference. The assignment of keys (“Z” and “/”) to the “same” and “different” responses was counterbalanced across participants. When a participant was incorrect, the word “Wrong” appeared on the screen for 2.2 s. During this time, if a pair actually had changed, the display flipped between the original and changed pair every 200 ms to highlight the difference the participant had failed to notice.

Half the trials were *different* trials, in which one pair changed, and the other half were *same* trials, with no changes. Half the *different* trials involved categorical differences, whereas the other half involved purely metric differences. On the *different* trials, the change appeared equally often in the four quadrants of the screen. Each trial randomly drew from either the set of circle pairs shown in Figure 3a or the mirror-reflected set. The noncritical circle pairs (the ones that did not change from the first to the second display) were randomly chosen from the same set, subject to the constraint that no display could contain two instances of the same circle pair.

Experiments 1a and 1b

In these experiments, we tested whether viewers were more accurate at detecting categorical changes than purely metric changes using set sizes of one and two circle pairs (Experiment 1a) and set sizes of two and three circle pairs (Experiment 1b; Fig. 3c). Separate samples of participants were recruited for the two experiments. Each experiment included a total of 256 trials: 2 trial types (same/different) \times 8 change intervals \times 2 directions of change in the smaller circle’s position \times 4 quadrants of change \times 2 set sizes. There were eight between-subjects conditions created by crossing the two response-key assignments with the four color schemes for the circles. Most of these conditions were tested twice in each experiment.

Experiment 1a

Fifteen Northwestern University students (11 female, 4 male) took part in this experiment for class credit. They received a self-timed break halfway through the experiment.

Figure 4 summarizes participants’ accuracy. We analyzed accuracy on *different* trials with a 2 (set size: one pair vs. two pairs) \times 2 (difference type: categorical vs. metric) repeated measures analysis of variance (ANOVA). Participants performed better on one-pair trials ($M = .914$) compared with two-pair trials ($M = .804$), $F(1, 11) = 31.6$, $p < .001$, $\eta_p^2 = .758$. Performance was also better for

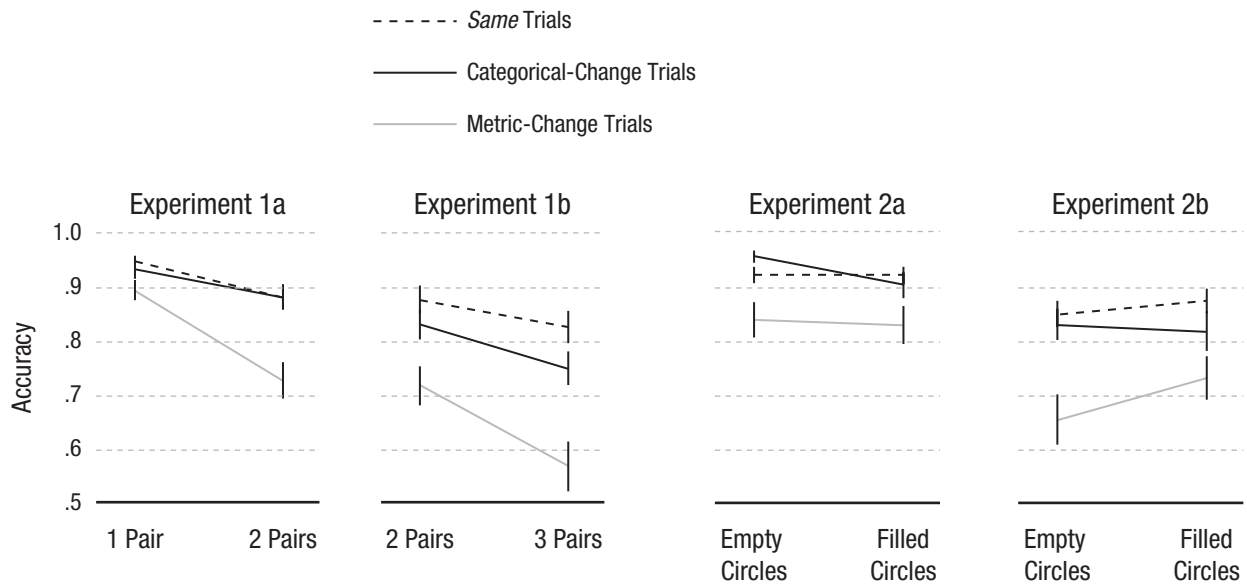


Fig. 4. Mean accuracy in the four experiments on *same* trials and *different* trials with categorical and metric changes. For Experiments 1a and 1b, accuracy is shown separately for one-pair and two-pair trials, and for Experiments 2a and 2b, accuracy is shown separately for trials with empty circles and trials with filled circles. Error bars represent ± 1 SE.

categorical-change trials ($M = .907$) compared with metric-change trials ($M = .810$), $F(1, 11) = 43.8$, $p < .001$, $\eta_p^2 = .693$. This categorical-change advantage was particularly apparent on two-pair trials, as evidenced by a significant interaction between set size and difference type, $F(1, 11) = 17.7$, $p = .001$, $\eta_p^2 = .559$. Because this interaction could have been driven by near-ceiling performance in the one-pair condition, in Experiment 1b we tested whether this interaction would be found when we compared two-pair with three-pair trials.

There was no evidence of a speed-accuracy trade-off. Although response speed was not emphasized in the task instructions, we analyzed response times for correct responses using the median time for each participant-condition pairing. Participants responded faster when there was only one circle pair ($M = 976$ ms) than when there were two ($M = 1,294$ ms), $F(1, 11) = 155.9$, $p < .001$, $\eta_p^2 = .918$, and they responded faster when the difference was categorical ($M = 1,105$ ms) than when it was metric ($M = 1,165$ ms), $F(1, 11) = 9.2$, $p = .009$, $\eta_p^2 = .395$. The interaction of set size and difference type was not significant, $p > .250$.

Experiment 1b

Fifteen participants, ages 18 through 35, took part in this experiment in return for \$10 (11 females, 4 males; mean age = 22.6 years). Participants received a self-timed break after every 64 trials.

Figure 4 summarizes participants' accuracy. We analyzed accuracy on *different* trials with a 2 (set size: two

vs. three pairs) \times 2 (difference type: categorical vs. metric) repeated measures ANOVA. Participants performed more accurately on two-pair trials ($M = .773$) compared with three-pair trials ($M = .657$), $F(1, 11) = 32.7$, $p < .001$, $\eta_p^2 = .700$. Performance was also more accurate for categorical-change trials ($M = .789$) compared with metric-change trials ($M = .642$), $F(1, 11) = 35.2$, $p < .001$, $\eta_p^2 = .715$. The categorical-change advantage was again greater on trials with the larger set size (three pairs), as evidenced by a significant interaction between set size and difference type, $F(1, 11) = 5.4$, $p = .036$, $\eta_p^2 = .278$, though the effect size of this interaction was considerably smaller in this experiment, which minimized the ceiling effect in the lower-set-size condition, than in Experiment 1a ($\eta_p^2 = .278$ vs. $.559$).

There was again no evidence of a speed-accuracy trade-off. We analyzed response times for correct responses using the median time for each participant-condition pairing. Participants responded faster when there were only two circle pairs ($M = 1,296$ ms) than when there were three ($M = 1,486$ ms), $F(1, 11) = 18.9$, $p = .001$, $\eta_p^2 = .574$. Neither the main effect of difference type ($p > .250$) nor the interaction ($p > .250$) was significant.

Accuracy as a function of interval

The analyses reported thus far collapsed across those differences we assumed to be categorical. But Figure 5, which depicts accuracy separately for each interval in those conditions with a set size of 2 or 3, suggests the need to revise the categories. To be sure, some

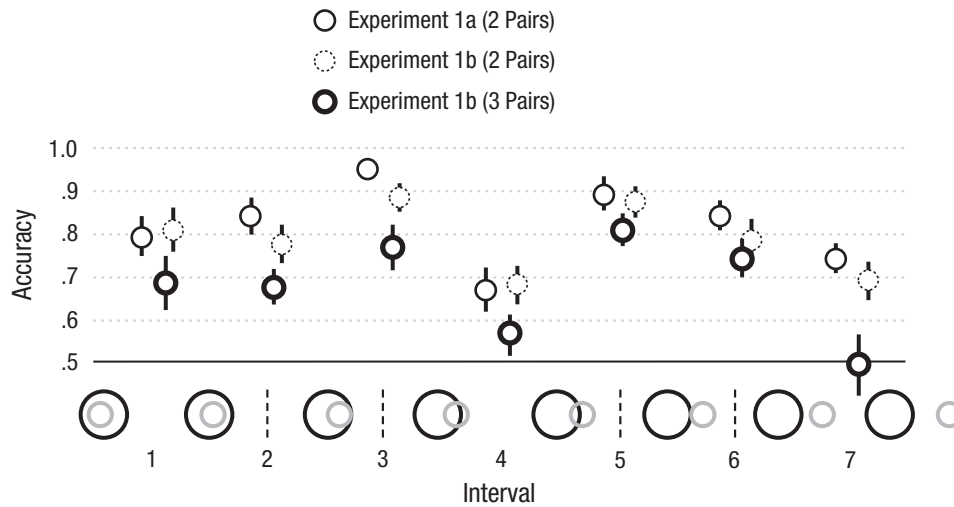


Fig. 5. Mean accuracy in the two- and three-pair trials of Experiments 1a and 1b as a function of interval. The diagrams along the x-axis show the eight circle pairs in the stimulus set; the dashed vertical lines indicate the intervals that were hypothesized to introduce a categorical change in the topological relations between the two circles in a pair. Error bars represent ± 1 SE.

results were as expected. Consider Intervals 4 and 7, which past modeling work predicted would be treated as purely metric changes. Indeed, participants' accuracy at detecting changes across these intervals was relatively low. We compared accuracy for these intervals with accuracy for Interval 2, the categorical interval with the lowest accuracy, via paired-samples *t* tests (Table 1). Accuracy was higher for Interval 2 than for Intervals 4 and 7 both with the set size of 2 and with the set size of 3, and the difference was significant or nearly significant in all cases.

But now consider Interval 1, which we hypothesized involved a purely metric change. Surprisingly, accuracy at detecting changes across this interval was comparable

to accuracy for putatively categorical changes. For example, in the three-pair condition of Experiment 1b, accuracy levels for Interval 1 ($M = .683$) were not significantly different from those for Interval 2 ($M = .675$), $p > .250$.

Discussion

In both experiments, changes in categorical relations were far easier to detect than purely metric changes. This effect was robust across changes in set size, and the advantage was moderately larger for larger set sizes, when memory load was higher, possibly because of a ceiling effect. However, differences across Interval 1,

Table 1. Pairwise Comparisons of Accuracy for Intervals 4 and 7 With Accuracy for Interval 2 in Experiments 1a and 1b

Experiment and condition	Metric interval	Mean accuracy	Comparison with Interval 2				
			Mean difference	95% CI	<i>t</i> (14)	<i>p</i>	<i>d</i>
Experiment 1a							
Set Size 2	7	.741	.101	[.008, .193]	2.33	.035	0.66
Set Size 2	4	.667	.175	[.083, .267]	4.07	.001	0.95
Experiment 1b							
Set Size 2	7	.690	.085	[-.008, .179]	1.96	.070	0.50
Set Size 2	4	.678	.097	[-.027, .220]	1.64	.116	0.56
Set Size 3	7	.493	.182	[.042, .322]	2.79	.014	0.81
Set Size 3	4	.565	.110	[-.023, .243]	1.77	.099	0.63

Note: Mean accuracy for Interval 2 was .842 in Experiment 1a, .775 for Set Size 2 in Experiment 1b, and .675 for Set Size 3 in Experiment 1b. The mean differences listed were calculated by subtracting the mean accuracies for Intervals 4 and 7 from these values. CI = confidence interval.

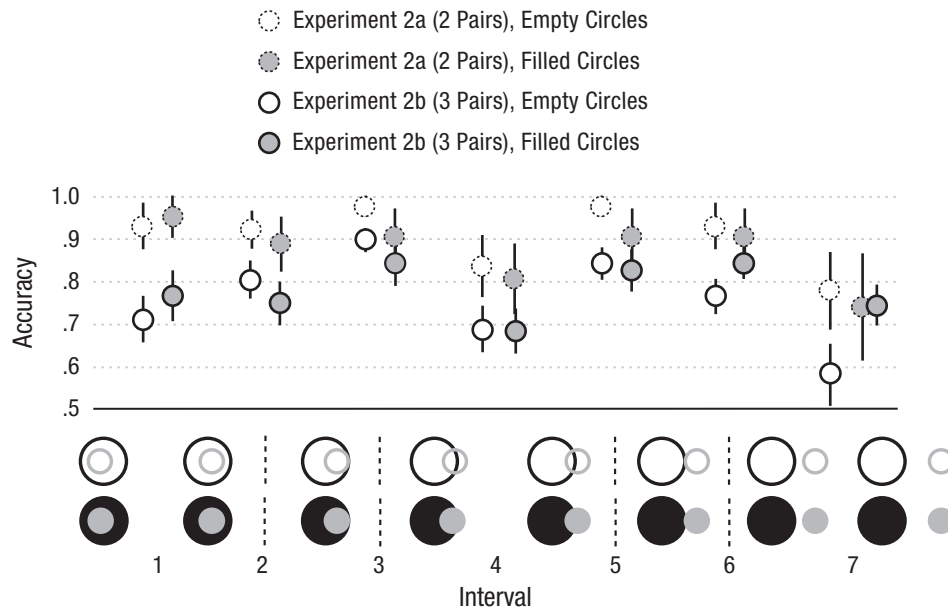


Fig. 6. Mean accuracy in empty-circle and filled-circle trials of Experiments 2a (two-pair trials) and 2b (three-pair trials) as a function of interval. The diagrams along the x-axis show the eight circle pairs in each stimulus set; the dashed vertical lines indicate the intervals that were hypothesized to introduce a categorical change in the topological relations between the two circles. Error bars represent $\pm 1 SE$.

one of our proposed metric intervals, appeared to be detected about as easily as the categorical differences. There may indeed be a perceived difference in category in Interval 1; that is, the smaller object may be coded as being to one side of the larger object's vertical midline and then the other, which would be consistent with evidence that people categorically code which quadrant of a circle contains a dot (Huttenlocher et al., 1991).

Experiments 2a and 2b

The categorical perception task used in Experiments 1a and 1b provided a paradigm for discovering the boundaries between relational categories, but also for exploring the cues that the visual system uses to construct those categories. For example, for Interval 5, the presence versus absence of the “overlapping” relation might be signaled by a change in the number of *enclosed regions* visible across the pair; overlapping increases the number of enclosed regions from two to three. Evidence from other tasks suggests that the number of enclosed regions may serve as a critical visual primitive in initial processing of a visual scene (Chen, 2005).

To provide a case study exploring the cues for relational categories, in Experiments 2a and 2b we tested whether the categories encoded depend on the number of enclosed regions, by equating the number of regions while manipulating the relational categories. We did this by introducing a condition with filled circles, so that there were always only two regions in a display

(see the illustrations below the graph in Fig. 6). If the number of enclosed regions is a critical cue, the categorical-change advantage would be weakened when the circles were filled.

In Experiment 2a, two circle pairs were shown in each display, whereas in Experiment 2b, three circle pairs were shown. Separate samples of participants were recruited for the two experiments. Each experiment included a total of 256 trials: 2 trial types (same/different) \times 8 change intervals \times 2 directions of change in the smaller circle's position \times 4 quadrants of change \times 2 circle types (filled/empty). The filled- and empty-circle trials were split into separate 128-trial blocks. The order of the two blocks varied across participants. The experiments were otherwise identical to Experiment 1b.

There were 16 between-subjects conditions created by crossing the two response-key assignments, the four color schemes for the circles, and the two block orders. Each condition was tested once in each experiment (with the exception that because of experimenter error, one condition was repeated and one condition was not run in Experiment 2a).

Experiment 2a

Sixteen participants, ages 18 through 35, took part in this study in return for \$10 (12 female, 4 male; mean age = 21.1 years). Participants' accuracy is summarized in Figure 4. We analyzed accuracy on *different* trials via a 2 (circle type: filled vs. empty) \times 2 (difference type:

Table 2. Analysis of Variance Results in Experiments 2a and 2b

Effect	Experiment 2a			Experiment 2b		
	$F(1, 14)$	p	η_p^2	$F(1, 14)$	p	η_p^2
Circle type	6.0	.028	.301	2.2	.160	.134
Difference type	14.8	.002	.515	35.4	< .001	.717
Block order	2.2	.163	.134	1.2	> .250	.079
Circle Type \times Difference Type	1.3	> .250	.085	5.5	.034	.282
Circle Type \times Block Order	22.9	< .001	.620	2.2	.163	.134
Difference Type \times Block Order	0.3	> .250	.018	3.5	.083	.199
Circle Type \times Difference Type \times Block Order	0.2	> .250	.015	0.3	> .250	.018

Note: Significant results are highlighted in boldface.

categorical vs. metric) \times 2 (block order: filled vs. empty circles first; between subjects) repeated measures ANOVA (see Table 2). There were significant main effects for circle type and difference type. Participants were more accurate with empty than with filled circles, and they were more accurate at detecting categorical differences than at detecting metric differences. There was also an interaction between block order and circle type; filled circles were harder than empty circles primarily when they were viewed first. No other effects were significant.

Pairwise comparisons confirmed that the categorical-change advantage was present for both filled and empty circles. For filled circles, mean accuracy for categorical changes was .902, and mean accuracy for metric changes was .822, mean difference = .080, 95% confidence interval (CI) = [.028, .132], $t(15) = 3.26$, $p = .005$, $d = 0.67$. For empty circles, mean accuracy for categorical changes was .951, and mean accuracy for metric changes was .838, mean difference = .113, 95% CI = [.047, .180], $t(15) = 3.63$, $p = .002$, $d = 1.15$. Note that these are conservative estimates of the categorical-change advantage, as Interval 1 was included with the metric changes.

Although the analyses presented thus far tested whether the effect of number of enclosed regions might have contributed to the categorical-change advantage, an alternative possibility is that differences in low-level visual features computed early in visual processing contribute to performance on this task. To test for this possibility, we computed the similarity between circles within each pair using a Gabor-jet model, which approximates responses in primary visual cortex (Yue, Biederman, Mangini, von der Malsburg, & Amir, 2012). Previous research suggests that when images lack significant categorical differences, the responses of this model correlate with human similarity judgments.

Each circle pair was reproduced as a gray-scale image, and the images were input into the model to generate low-level feature vectors. Our measure of difference was the euclidean distance between feature

vectors. For the empty circles, the correlation between modeled differences and participants' accuracy was positive but nonsignificant, $r(5) = .36$, $p > .250$. For the filled circles, the correlation again was positive but nonsignificant, $r(5) = .50$, $p > .250$. These results suggest that performance was not driven primarily by low-level image differences. However, because the results are tied to the particular similarity model used, we cannot rule out the possibility that untested visual features contributed to participants' discrimination performance.

Experiment 2b

Sixteen participants, ages 18 through 35, took part in this study in return for \$10 (11 female, 5 male; mean age = 21.7 years). Participants' accuracy is summarized in Figure 4. We analyzed accuracy on *different* trials via a 2 (circle type: filled vs. empty) \times 2 (difference type: categorical vs. metric) \times 2 (block order: filled vs. empty circles first; between subjects) repeated measures ANOVA (see Table 2). There was a main effect for difference type: Participants were more accurate at detecting categorical than metric differences. There was also an interaction between difference type and circle type: The categorical-change advantage decreased with filled circles. Despite the interaction, pairwise comparisons confirmed that the categorical-change advantage was present for both filled and empty circles. For filled circles, mean accuracy for categorical changes was .816, and mean accuracy for metric changes was .730, mean difference = .086, 95% CI = [.040, .132], $t(15) = 3.97$, $p = .001$, $d = 0.56$. For empty circles, mean accuracy for categorical changes was .828, and mean accuracy for metric changes was .652, mean difference = .176, 95% CI = [.097, .254], $t(15) = 4.77$, $p < .001$, $d = 1.14$.

Accuracy as a function of interval

The reduction in the categorical-change advantage for filled circles relative to empty circles appears to have been driven by Interval 7 (see Fig. 6). In Experiment

2b, participants were far better at detecting differences across this metric change when the circles were filled ($M = .743$) than when they were empty ($M = .581$), $p = .015$. This effect was not present in Experiment 2a, in which only two circle pairs were used, so it appears to have been driven by the greater number of pairs in Experiment 2b. We conducted an additional ANOVA that collapsed across Experiments 2a and 2b, including number of pairs as a between-subjects variable. This ANOVA confirmed the reported categorical advantage and also produced a significant interaction between number of pairs and circle type, $F(1, 16) = 5.6$, $p = .025$, $\eta_p^2 = .167$. With two pairs, accuracy was slightly higher for empty circles, but with three pairs, accuracy was slightly higher for filled circles, a difference we attribute to Interval 7. In addition, as expected there was a main effect for number of pairs; accuracy was higher for two pairs than for three, $F(1, 16) = 10.1$, $p = .004$, $\eta_p^2 = .265$.

Discussion

Experiments 2a and 2b both replicated the advantage for detecting changes that cross categorical boundaries and showed that this advantage is robust for both empty and filled circles. However, in Experiment 2b, the categorical-change advantage was weakened for filled circles because of accurate performance across metric Interval 7 when the circles were filled (Fig. 6). One possible reason for this difference between Experiments 2a and 2b is that filling the circles increased representation of the objects at low spatial frequencies, which may have provided a better cue for representing the white space between them, or the aspect ratio of the envelope that surrounded them (Badcock, Whitworth, & Badcock, 1990; Thomas, Kveraga, Huberle, Karnath, & Bar, 2012), thereby aiding change detection across Interval 7. A low-resolution representation capturing the overall arrangement of the objects may have been used more when the set size was larger (three vs. two circle pairs) to address the greater working memory load.

Conclusion

Our participants remembered and compared images by relying on a set of categorical relations between objects. This effect was robust across changes in set size, and there was evidence that participants recruited additional categorical relations to aid in their judgments (“left” vs. “right” within the large circle in the case of Interval 1). We also found evidence against the hypothesis that topological relations are encoded on the basis of the number of enclosed regions. Thus, our results leave open the question of what visual cues drive the encoding of topological relations, while hinting that the

overall arrangement of objects may play a role in some cases.

Our data do not make it possible to isolate whether the set of categorical relations we tested brings the strongest benefit at the encoding, retrieval, or comparison stages of processing, although the computational models (Lovett & Forbus, 2011a, 2017; Lovett et al., 2009) posit the strongest role during comparison. Our data also cannot conclusively answer the question of how strongly the visual categories are linked to lexical representations (e.g., words such as *overlapping*). Although there is evidence that the categorical perception of colors interacts with verbal coding (Roberson, Pak, & Hanley, 2008; Winawer et al., 2008), other evidence suggests that the type of spatial category advantage we studied is tied more closely to perceptual processing (Kranjec et al., 2014), and even particular visual processing areas (Kim & Biederman, 2012). Regarding this question, the computational models can say little because none of the modeled tasks included a language component. Future work might incorporate verbal interference manipulations to investigate whether certain relations (e.g., the “left” vs. “right” distinction, which is known to develop more slowly than other relations; Rigal, 1994) rely more heavily than others on verbal coding. Another area ripe for study would be cross-linguistic (e.g., English vs. Korean) differences in relational categorizations; for example, the distinction between “in” and “on” could be tested in different language populations with our categorical perception task, perhaps using more realistic 3-D stimuli (e.g., Bowerman, 1996).

This experiment should serve as a template for future research exploring the suite of categories people perceive when they view between-object relations. Future research could further explore the cues used by the visual system to detect such categories, through systematic variation of low-level object properties, as in the empty-versus-filled manipulation of Experiments 2a and 2b. For example, one could examine how strongly topological relations are driven by the presence and arrangement of edge junctions, such as an X-junction where the edges of two empty circles overlap and a T-junction where one filled circle occludes another (Cavanagh, 1987; Peterson & Hochberg, 1983). This template can be used more broadly to evaluate the set of categorical relations proposed in modeling work (e.g., Lovett & Forbus, 2011b), including relations for relative position (*above*, *right of*), alignment (*parallel*, *collinear*), and shape transformations (*rotation between*, *reflection between*).

Ultimately, a better understanding of these categorical relations will help explain how the human visual system compares structure across diagrams and graphs

used by students and scientists, while also providing concrete guidelines for designers who wish to facilitate comparison by crossing categorical boundaries.

Action Editor

Philippe G. Schyns served as action editor for this article.

Author Contributions

A. Lovett developed the study concept. Both authors contributed to the study design. Testing, data collection, and data analysis were performed by A. Lovett under the supervision of S. L. Franconeri. Both authors contributed to drafting and revising the manuscript, and both authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by a National Institutes of Health training grant in human cognition at Northwestern University (T32 NS047987), as well as by National Science Foundation Grants BCS-1056730 CAREER and IIS-1162067.

Open Practices



All data have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/fhegd/>. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797617709814>. This article has received the badge for Open Data. More information about the Open Practices badges can be found at <https://www.psychologicalscience.org/publications/badges>.

References

- Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (1998). *Essential cell biology: An introduction to the molecular biology of the cell*. New York, NY: Garland.
- Amir, O., Biederman, I., Herald, S. B., Shah, M. P., & Mintz, T. H. (2014). Greater sensitivity to nonaccidental than metric shape properties in preschool children. *Vision Research*, *97*, 83–88.
- Badcock, J. C., Whitworth, F. A., & Badcock, D. R. (1990). Low-frequency filtering and the processing of local-global stimuli. *Perception*, *19*, 617–629.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.
- Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research*, *46*, 207–222.
- Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 493–530). Cambridge, MA: MIT Press.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*, 404–431.
- Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing*, *37*, 171–195.
- Chen, L. (2005). The topological approach to perceptual organization. *Visual Cognition*, *12*, 553–637.
- Cirillo, S., & Ström, V. (2010). *An anthropomorphic solver for Raven's Progressive Matrices* (Chalmers University of Technology, Department of Applied Information Technology, Report No. 2010:096). Retrieved from http://www.sais.se/mthprize/2011/cirillo_strom.pdf
- Doumas, L. A. A., & Hummel, J. E. (2013). Comparison and mapping facilitate relation discovery and predication. *PLoS ONE*, *8*(6), Article e63889. doi:10.1371/journal.pone.0063889
- Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*, 227–240.
- Falkenhainer, B., Forbus, K., & Gentner, D. (1989). The structure mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*, 1–63.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.
- Hummel, J. E., & Stankiewicz, B. J. (1996). Categorical relations in shape perception. *Spatial Vision*, *10*, 201–236.
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*, 352–376.
- Kim, J. G., & Biederman, I. (2012). Greater sensitivity to non-accidental than metric changes in the relations between simple shapes in the lateral occipital cortex. *NeuroImage*, *63*, 1818–1826.
- Klippel, A., Yang, J., Wallgrün, J. O., Dylla, F., & Li, R. (2012). Assessing similarities of qualitative spatiotemporal relations. In C. Stachniss, K. Schill, & D. H. Uttal (Eds.), *Spatial cognition 2012* (pp. 242–261). Berlin, Germany: Springer.
- Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. (1977). Category and continuum in mental comparisons. *Journal of Experimental Psychology: General*, *106*, 341–375.
- Kranjec, A., Lupyan, G., & Chatterjee, A. (2014). Categorical biases in perceiving spatial relations. *PLoS ONE*, *9*(5), Article e98604. doi:10.1371/journal.pone.0098604
- Kunda, M., McGreggor, K., & Goel, A. K. (2013). A computational model for solving problems from the Raven's

- Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research*, 22–23, 47–66.
- Larkey, L., & Love, B. (2003). CAB: Connectionist Analogy Builder. *Cognitive Science*, 27, 781–794.
- Lovett, A., & Forbus, K. (2011a). Cultural commonalities and differences in spatial problem-solving: A computational analysis. *Cognition*, 121, 281–287.
- Lovett, A., & Forbus, K. (2011b). Organizing and representing space for visual problem-solving. In *Proceedings of the 25th International Workshop on Qualitative Reasoning (QR'11)*. Retrieved from http://www.qrg.northwestern.edu/papers/Files/qg-workshops/QR2011/QR11_Proceedings_Index.htm
- Lovett, A., & Forbus, K. (2017). Modeling visual problem-solving as analogical reasoning. *Psychological Review*, 124, 60–90.
- Lovett, A., Tomai, E., Forbus, K., & Usher, J. (2009). Solving geometric analogy problems through two-stage analogical mapping. *Cognitive Science*, 33, 1192–1231.
- MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 369–390.
- Peterson, M. A., & Hochberg, J. (1983). Opposed-set measurement procedure: A quantitative analysis of the roles of local cues and intention in form perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 183–193.
- Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A spatial logic based on regions and connection. In B. Nebel, C. Rich, & W. Swartout (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the Third International Conference* (pp. 165–176). San Mateo, CA: Morgan Kaufmann.
- Rigal, R. (1994). Right-left orientation: Development of correct use of right and left terms. *Perceptual & Motor Skills*, 79, 1259–1278.
- Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, 107, 752–762.
- Rosielle, L. J., & Cooper, E. E. (2001). Categorical perception of relative orientation in visual object recognition. *Memory & Cognition*, 29, 68–82.
- Thomas, C., Kveraga, K., Huberle, E., Karnath, H. O., & Bar, M. (2012). Enabling global processing in simultanagnosia by psychophysical biasing of visual pathways. *Brain*, 135, 1578–1585.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2008). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, USA*, 104, 7780–7785.
- Yue, X., Biederman, I., Mangini, M. C., von der Malsburg, C., & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. *Vision Research*, 55, 41–46.