

Evolutionary Constraints on Human Object Perception

Sarah E. Koopman,^a Bradford Z. Mahon,^{a,b,c,d} Jessica F. Cantlon^a

^a*Department of Brain & Cognitive Sciences, University of Rochester*

^b*Department of Neurosurgery, University of Rochester Medical School*

^c*Center for Visual Science, University of Rochester*

^d*Center for Language Science, University of Rochester*

Received 4 March 2016; received in revised form 16 September 2016; accepted 10 October 2016

Abstract

Language and culture endow humans with access to conceptual information that far exceeds any which could be accessed by a non-human animal. Yet, it is possible that, even without language or specific experiences, non-human animals represent and infer some aspects of similarity relations between objects in the same way as humans. Here, we show that monkeys' discrimination sensitivity when identifying images of animals is predicted by established measures of semantic similarity derived from human conceptual judgments. We used metrics from computer vision and computational neuroscience to show that monkeys' and humans' performance cannot be explained by low-level visual similarity alone. The results demonstrate that at least some of the underlying structure of object representations in humans is shared with non-human primates, at an abstract level that extends beyond low-level visual similarity. Because the monkeys had no experience with the objects we tested, the results suggest that monkeys and humans share a primitive representation of object similarity that is independent of formal knowledge and cultural experience, and likely derived from common evolutionary constraints on object representation.

Keywords: Object representation; Evolution; Non-human primate; Homology; Representational structure

1. Introduction

Human object knowledge is structured by *semantic similarity*—the relations among objects are defined by their perceptual and conceptual properties such as form, function, behavior, and environment (e.g., Collins & Loftus, 1975; Cree & McRae, 2003; Martin, 2007, 2009; Warrington & McCarthy, 1987). Some of the conceptual properties that humans

Correspondence should be sent to Sarah E. Koopman, Department of Brain & Cognitive Sciences, University of Rochester, Meliora Hall Box 270268, Rochester, NY 14627. E-mail: skoopman@ur.rochester.edu

represent about objects are observable (e.g., has a beak), whereas others are inferred (e.g., is alive); relatedly, some conceptual properties can be observed directly and represented non-verbally (e.g., flies), whereas others are most often acquired indirectly through verbal transmissions (e.g., “lives in Antarctica”). Semantic information also can be inferred from perceptual features—humans can combine perception and prior knowledge to make predictions about the likely functional, behavioral, and thematic properties of a novel object (Carey, 1982; Keil, 1981). Semantic representations are useful because they allow humans to make inferences about novel objects that they have never experienced. The origin of human semantic structure, and its possible evolutionary foundation in primate cognition, is a current focus of enquiry in child development, psychology, and neuroscience (Humphreys, Price, & Riddoch, 1999; Keil, 1981; Markman, 1992; Sloutsky & Fisher, 2004).

Representations of objects could be similar between humans and non-human primates only superficially, in terms of the low-level perceptual representations, or they could also be similar at more abstract perceptual levels of representation. The question of whether humans and non-human primates have similar representations of similarity relations between objects is important for understanding which aspects of human object representations are evolutionarily derived, and which aspects critically depend on human-specific abilities, such as culture and language. Human conceptual reasoning far exceeds the conceptual reasoning abilities of non-human animals (Penn, Holyoak, & Povinelli, 2008). But, as in other domains, such as numerical cognition (Brannon & Terrace, 1998; Cantlon & Brannon, 2006; Gallistel & Gelman, 1992) and social cognition (Spelke, Bernier, & Skerry, 2013; Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012), it could be the case that a common set of predictable attributes underlies the basic structure of object knowledge in humans and non-human primates (Keil, 1981; New, Cosmides, & Tooby, 2007; Phillips & Santos, 2007).

Cognitive research with non-human animals indicates that they can form basic categories. For instance, a variety of non-human animal species can be trained to sort visual images into categories, such as trees, water, leaf patterns, particular people, flowers, cars, and even painting styles (Cerella, 1979; Herrnstein & Loveland, 1964; Herrnstein, Loveland, & Cable, 1976; Wasserman, Kiedinger, & Bhatt, 1988; Watanabe, Sakamoto, & Wakita, 1995). Some aspects of non-human animals’ categorization abilities have human-like qualities, including discrimination at the superordinate and basic levels (Bovet & Vauclair, 2001; Fabre-Thorpe, Richard, & Thorpe, 1998; Sands, Lincoln, & Wright, 1982; Vogels, 1999), with rapid category discrimination and generalization to novel exemplars (Fabre-Thorpe et al., 1998; Neiworth & Wright, 1994; Vauclair & Fagot, 1996). Additionally, non-human primates exhibit some signatures of abstract representation when identifying objects according to their superordinate categories. For example, macaque monkeys accurately categorize animals (vs. non-animals) and show similar congruency effects to humans when an object (animal or man-made) matches its background (natural or man-made) (Fize, Cauchoux, & Fabre-Thorpe, 2011). However, the degree to which the similarity space of object representations is related to that of non-human primates remains unknown.

One reason to propose that human object knowledge shares structural attributes with non-human primate cognition is that there are numerous homologies between the neural

systems of humans and non-human primates due to millions of years of shared evolutionary history. In the visual system, anatomical and functional similarities in the human and non-human primate brain extend from early visual regions through to high-level visual areas that process whole objects (Kravitz, Saleem, Baker, & Mishkin, 2011; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013; Kriegeskorte, Mur, Ruff, et al., 2008; Logothetis, Guggenberger, Peled, & Pauls, 1999; Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Van Essen & Gallant, 1994; Wandell, 1999). A growing body of electrophysiological and neuroimaging research provides further evidence of similar neural representations of objects in monkeys and humans (Gil-da-Costa et al., 2004). Kriegeskorte, Mur, Ruff, et al. (2008) directly compared monkey and human inferior temporal (IT) responses to the same visual object stimuli and found that the patterns of neural responses in humans and monkeys could be used to group the stimuli into similarly organized taxonomic structures. However, despite evidence of neural homologies between humans and non-human primates, the question of whether humans and non-human primates represent information similarly during judgments about objects is debated (Fize et al., 2011; Gil-da-Costa et al., 2004; Mur et al., 2013).

In order to bridge existing neurobiological research on monkeys' representations of visual objects with cognitive science research on human cognition, we need behavioral evidence that monkeys' cognitive judgments of objects are predicted by and share representational structure with human representations. One way to directly test this is to use human-derived measures of the semantic distances among objects to test whether those human-based measures predict monkeys' discrimination performance. It is known that human-derived measures of semantic distance among objects reflect both abstract perceptual and formal knowledge, and critically, depend on humans' rich knowledge of the abstract properties of objects, including thematic, behavioral, and functional properties (Cree & McRae, 2003). Although previous research demonstrates that animals can learn object categories and discriminate objects independently of global visual scene statistics, it does not show whether animals possess representations of categories with similar semantic structure to abstract human categories.

Here, we directly compared humans' and monkeys' representations of object similarity for the category "animals" in a non-verbal matching task. We found patterns of similarity between humans and monkey representational spaces. In order to identify the sources of those similarities between monkeys' and humans' representations, we compared metrics of human semantic similarity with several metrics of visual similarity to test explanations of animals' discrimination sensitivity, including metrics from a computational model of low-level visual processing in visual cortex (V1) and higher level whole object processing in IT cortex. If monkeys' representations of objects are predominantly driven by simple visual representations, then their discriminations will be predicted by metrics of low-level visual similarity alone. However, if monkeys' representations of objects are more abstract, then their discriminations will be predicted by metrics of IT representation and human semantic similarity indices, over and above any contributions of visual similarity. Evidence for abstract conceptual representations in monkeys would implicate a role for evolutionary constraints on the organization of category representations, at a higher level of abstraction than has been demonstrated previously.

2. Experimental procedures

2.1. Match-to-sample task—Line drawings

This task tested the representational similarity space of object representations, using *line drawings* as stimuli, in monkeys and humans.

2.1.1. Subjects

Monkey subjects were four rhesus macaques (*Macaca mulatta*; two female). Two monkeys (D and R) were water-restricted and received juice rewards in the task. The other two monkeys (C and H) had access to water *ad libitum* and received treat pellets and M&M rewards in the task. All monkeys were familiar with the match-to-sample paradigm, but not with the stimuli used in this experiment. Each monkey completed 12 sessions of 90 trials each (approximately 45 min per session). All procedures were approved by the University of Rochester IACUC.

Human participants were 10 undergraduates (8 female) recruited from the University of Rochester. All participants gave written informed consent and were compensated \$8/h for their participation. Each participant completed 1 session of 450 trials (approximately 45 min). All procedures were approved by the University of Rochester Research Subjects Review Board (RSRB).

2.1.2. Stimuli

Forty-five unique pairs of exemplars from the category “animals” were tested (i.e., all possible pairings of 10 objects). The monkeys had never seen these stimuli before, and the data presented here represent their entire experience with these stimuli. The same black-and-white line drawings of animals were used in the match-to-sample task and in the picture–word agreement task (Fig. 1A).

2.1.3. Procedure

The match-to-sample task was conducted using touchscreen computers and was identical for monkeys and humans (and for line drawings, photographs, and novel objects—see below). Each trial began with a green screen with a white rectangle in the lower right-hand corner. To initiate a trial, subjects touched the white rectangle. After a 1-s delay, a picture stimulus appeared randomly at one of six positions on the screen. The subject touched the stimulus to verify s/he was attending to the task. The stimulus disappeared and, after a 2-s delay, two picture stimuli appeared randomly at two of six possible positions on the screen. One of these stimuli was a match to the previously presented stimulus, while the other was not. If the subject touched the matching stimulus, a chime sound played, the screen turned pink for the 2-s inter-trial interval, and juice or pellets were delivered (monkeys only). If the subject touched the non-matching stimulus, or failed to respond in 8 s, a buzzer sounded and the screen turned black for a 5-s timeout before the initiation of the 2-s inter-trial interval.

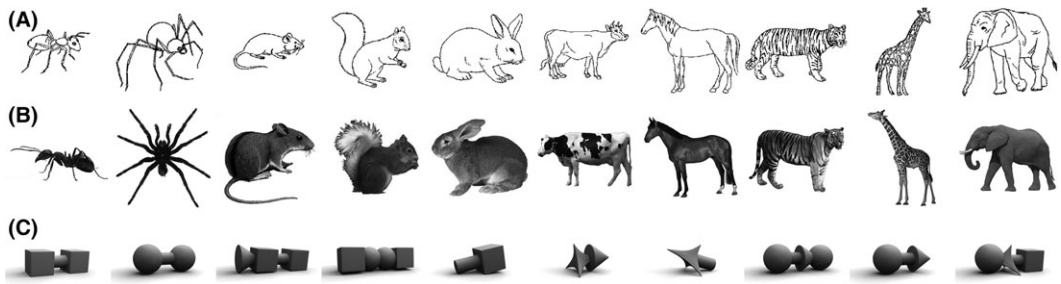


Fig. 1. Picture stimuli. (A) Black-and-white line drawings used in the animal line drawing match-to-sample task and the picture–word agreement task. (B) Grayscale photos used in the animal photograph match-to-sample task. (C) Grayscale photos of the geometric stimuli.

2.2. Match-to-sample task—Grayscale photographs

This task tested the representational similarity space of object representations using *photographs* as stimuli, in monkeys and humans.

2.2.1. Subjects

Monkey subjects were two rhesus macaques (*Macaca mulatta*; 1 female), both of which had previously participated in the match-to-sample task with line drawings (R and H). Each monkey completed 12 sessions of 90 trials each (approximately 45 min per session).

Human participants were 12 undergraduates (8 female) recruited from the University of Rochester, who had not participated in the task with animal line drawings. All participants gave written informed consent and were compensated \$8/h for their participation. Each participant completed one session of 450 trials (approximately 45 min).

2.2.2. Stimuli

Stimuli were grayscale photographs of the same animal concepts tested with line drawings; 45 unique pairs of 10 animals were tested with this novel image set (Fig. 1B).

2.2.3. Procedure

All aspects of the procedure with this novel image set were exactly the same as in the experiment with line drawings.

2.3. Match-to-sample task—Novel geometric objects

This task tested the representational similarity space of *simple geometric objects*, using photographs as stimuli, in monkeys only.

2.3.1. Subjects

Monkey subjects were two rhesus macaques (*Macaca mulatta*; 1 female), both of which had previously participated in the match-to-sample task with animal line drawings

and animal photographs (R and H). Each monkey completed 12 sessions of 90 trials each (approximately 45 min per session).

2.3.2. *Stimuli*

Stimuli were 10 grayscale images of geometric objects composed of two to four parts, where each part was a simple, primitive shape such as a cube or cylinder (Fig. 1C). Forty-five unique pairs of 10 geometric objects were tested.

2.3.3. *Procedure*

All aspects of the procedure were exactly the same as in the experiments with line drawings and grayscale photographs, with the only difference that images of novel objects were used as stimuli.

2.4. *Picture–word agreement task*

This task tested the representational similarity space of object representations, using *words* as stimuli, in humans only.

2.4.1. *Subjects*

Human participants were 17 undergraduates (12 female) recruited from the University of Rochester. All participants gave written informed consent and were compensated \$8/h for their participation. Each participant completed 1,800 trials (approximately 3 h).

2.4.2. *Stimuli*

The same 45 unique pairings among the 10 animal concepts were tested as were tested in the Match-to-Sample tasks, and the same black-and-white line drawings of animals were used as had been used in the Match-to-Sample task (Fig. 1A). Word stimuli were the corresponding animal names in all capital letters and white font. Half of the trials were “yes” (i.e., match) trials and half of the trials were “no” (i.e., non-match) trials. Only response times and accuracy from “no” trials were analyzed, as the semantic distance in the match trials was always zero.

2.4.3. *Procedure*

The picture–word agreement task was conducted using a computer monitor and a response box. On each trial, a picture stimulus appeared in the middle of a black screen, with a word below it. The participant was instructed to press the left button on the response box if the picture and word represented the same concept, and the right button if the picture and word represented different concepts.

2.5. *Analysis*

Individual differences in speed-accuracy tradeoffs can affect whether semantic distance effects appear in subjects’ accuracy or in their response time. There can be large individual

differences in speed-accuracy tradeoffs among humans (Phillips & Rabbit, 1995), and even larger differences between primate species. For example, on discrimination tasks monkeys tend to respond rapidly with a moderate amount of error, whereas humans respond slowly with few errors (e.g., Cantlon & Brannon, 2006). Since semantic distance effects can be manifested in both response time and accuracy, we combined those two measures into one measure, which we refer to as “comparison cost.” This comparison cost measure, also called “inverse efficiency” (Townsend & Ashby, 1978, 1983), has been used in previous studies of semantic processing and discrimination (Aydelott, Baer-Henney, Trzaskowski, Leech, & Dick, 2012; Wei et al., 2012) and removes confounding individual differences in speed-accuracy tradeoff from analyses. First, for each subject (human or monkey), we calculated the median response time on correct trials for each image pair (or picture–word pair in the picture–word agreement task), after excluding outliers ($\pm 2 SD$; 4% of correct trials). We then calculated the mean of the medians for all subjects, for each image pair. Finally, we divided the mean response time for each stimulus pair by the overall accuracy for that image pair, yielding a comparison cost score. Thus, higher “comparison costs” correspond to longer response times and lower accuracy.

We validated the comparison cost measure by comparing monkey and human performance using two additional measures. First, we calculated the Spearman correlation between monkey accuracy and human response time. Second, we converted response time and error rate ($1 - \text{accuracy}$) to z -scores and then averaged them. We then calculated the Spearman correlation between the means of the z -scored response time and accuracy for monkeys and humans.

To support averaging data across individual monkeys, we calculated the internal consistency of subjects’ performance using *omega* (McDonald, 1999). Omega was 0.74 across individual monkey comparison costs for the line drawing and photo tasks, an acceptable level of internal consistency for averaging the data.

2.6. Semantic similarity measure

Semantic distance was defined using classic indices from human research on conceptual knowledge (Cree & McRae, 2003; McRae, Cree, Seidenberg, & McNorgan, 2005). McRae et al. (2005) quantified semantic distances among a large set ($n = 541$) of concepts by asking human subjects to list properties for each object. Subjects were encouraged to list a variety of types of properties, including perceptual, functional, and encyclopedic. The production frequency of every property generated by the subjects was calculated for each concept, creating a vector of production frequencies for each concept. Semantic distances among the concepts were then calculated using the cosine of property production frequency vectors in a pairwise fashion. Substantial research indicates that these ratings accurately reflect abstract human concepts (Maki & Buchanan, 2008; Mirman, 2011; Mirman & Magnuson, 2008, 2009) and capture substantial variance in response times during both verbal and non-verbal tasks (e.g., Caramazza, Hersh, & Torgerson, 1976; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Rips, Shoben, & Smith, 1973; Vigliocco, Vinson, Lewis, & Garrett, 2004).

To confirm that this property list-based measure reflects the semantic relationships between the animal stimuli, we compared it to explicit semantic similarity judgments obtained from adult participants. We asked 16 human participants to rate the pairwise similarity between the animals used in our experiments (presented as words) on a 7-point Likert scale, with 1 being “not similar at all” and 7 being “extremely similar.” We then used the median score for each pair (as recommended by Jamieson, 2004) in further analyses. This measure of semantic similarity was significantly correlated with the property list-based measure of semantic similarity from Cree & McRae (Spearman’s $r = .71$, $p < .01$), supporting our use of the property-based measure.

2.7. Visual similarity measures

Correlation matrices of all the visual similarity measures for animal line drawings (Table S1), animal photos (Table S2), and novel objects (Table S3) are included in the Supporting Information.

2.7.1. Affine similarity

This measure quantifies visual similarity by calculating how much one image must be transformed in order to match a second image. It is based on Belongie et al.’s (2002) “affine cost” measure, which is a measure of the difference between 100 arbitrarily chosen points on one image and “corresponding” points (as determined by an affine plane transformation model) on another image. The cost is the Euclidean distance between the points on the first image and the points on the second image (Belongie et al., 2002). This was carried out 100 times for each pairing of images, and the average cost (across the 100 samplings) was used to represent the affine cost. The signs of these values were changed (values were multiplied by -1) so that higher values correspond to greater similarity, hence “affine similarity.”

2.7.2. Shape context similarity

This measure is similar to affine similarity, with the addition that the context of each point was taken into account when choosing the “corresponding” point (shape context cost; Belongie et al., 2002). This was carried out 100 times for each pairing of images, and the average cost (across the 100 samplings) was used to represent the shape context cost. The sign of these values were changed (values were multiplied by -1) so that higher values correspond to greater similarity. Since it was unknown whether the addition of the shape context would improve the model’s representation of the visual similarity of our stimuli, we included both the affine and shape context similarity measures in our analyses.

2.7.3. Pixel correlation

This visual similarity measure was calculated by simply finding the Pearson linear correlation between pixel values of two images (e.g., Kriegeskorte, Mur, Ruff, et al., 2008).

2.7.4. *Sum of squared differences in pixel values-inverted*

The sum of squared differences in pixel values is frequently used for object matching in computer vision (e.g., Crowley, 1997; Grill-Spector, Kourtzi, & Kanwisher, 2001; Kanade & Okutomi, 1994). The differences between the values of the pixels in one image and the values of the corresponding pixels in a second image were calculated and squared. Then all the values were added to yield one value. The sign of this value was multiplied by -1 so that higher values correspond to greater similarity, hence “inverted.”

2.7.5. *Binary pixel correlation*

For the animal photo and novel object stimuli, the grayscale pixel values were converted into binary values using Otsu’s method (Otsu, 1979), and then the Pearson linear correlation between these pixel values of two images was found.

2.7.6. *Binary sum of squared differences in pixel values-inverted*

For the animal photo and novel object stimuli, the binary pixel values found using Otsu’s method (Otsu, 1979), were used to calculate the sum of squared differences in pixel values-inverted using the method described above.

2.7.7. *Correlation of radon transformed images*

Wade and Tyler (2005; see Kriegeskorte, Mur, & Bandettini, 2008, for details) suggested that the representation of an image in the lateral occipital complex (which is implicated in the perception and representation of objects (Snodgrass & Vanderwart, 1980)) can be modeled by the radon transformation of that image. This measure was calculated by Pearson correlating the radon transform matrices of two images (found using the radon function in MATLAB; e.g., Mur et al., 2013).

2.7.8. *V1(HMAX-S1C1) and IT (HMAX-C2)*

The V1 (HMAX-S1C1) and IT (HMAX-C2) visual similarity measures were calculated, using the HMAX model (Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; Serre, Wolf, & Poggio, 2005), which is a hierarchical, feed-forward neural network model of object recognition in the visual system. The first two layers of HMAX simulate simple and complex cells of primary visual cortex. The simple units’ (S1) responses are determined by passing the input image through Gabor filters. These responses serve as inputs to the complex units (C1), which respond to certain orientations of edges and are moderately tolerant to the shift and size of the inputs. The outputs of these layers are then put through a second set of simple and complex units (S2 and C2), which model the representation of the stimuli in IT cortex in terms of sets of features. S2 compares patches of the outputs of C1 to prototype patches, which have been determined by processing a set of training images through the C1 level. Finally, C2 extracts features over various positions and sizes, modeling the shift- and size-tolerant cells in IT. There is disagreement about the extent to which the C2 units represent IT cells (e.g., Khaligh-Razavi & Kriegeskorte, 2014); we therefore frequently refer to these visual similarity measures by both the model layers and the brain regions these layers purport to represent.

We used the model parameters described in Serre, Oliva, et al. (2007) and Serre et al. (2005) and the prototype patches trained from a set of natural images (e.g., Kriegeskorte, Mur, Ruff, et al., 2008). The prototype patches and HMAX model were downloaded from the Center for Biological & Computational Learning website (<http://cbcl.mit.edu/software-datasets/standardmodel/index.html>).

To calculate the V1 (HMAX-S1C1) visual similarity measure, the outputs of S1 and C1 were reshaped and concatenated to form one long vector for each image (e.g., Kriegeskorte, Mur, Ruff, et al., 2008). We then calculated the pairwise Pearson correlations between vectors, yielding a V1 visual similarity measure for each pair of stimuli. The IT (HMAX-C2) visual similarity measure was calculated in the same way as the V1 measure, using the outputs of C2. These metrics allow us to test the role of low-level visual feature similarity (V1) and whole object, abstract perceptual similarity (IT) in the monkeys' and humans' judgments.

3. Results

3.1. Match-to-sample task with animal stimuli

Monkeys and humans completed a match-to-sample task in which they discriminated images of animals. The task required subjects to choose a stimulus from two options that matched a sample stimulus that had just been presented. One choice option was an identical match and the other was a within-category foil (i.e., another animal). The semantic distance between the foil and the target was varied across trials. We used 10 animals (i.e., 45 unique pairs) from the larger database of Cree and McRae (2003) in this study.

In order to familiarize monkeys with the task, monkeys were trained to perform a match-to-sample task with an independent set of simple shapes and images. Once monkeys understood the matching task (>75% accuracy), they were tested on the critical stimuli. Subjects had no experience with the images used in this study prior to this experiment. In the first study, monkeys and humans performed the match-to-sample task over black-and-white line drawings of animals (Snodgrass & Vanderwart, 1980), as shown in Fig. 1A. Humans and monkeys performed significantly above chance on the task ($p < .01$; Monkey: 75%, $SD = 2\%$; Human: 99%, $SD = 0.5\%$) and with rapid response times (Median RTs: Monkey = 700 ms, $SD = 365$ ms; Human = 583 ms, $SD = 249$ ms). Because more variability in performance was captured by accuracy than response time in monkeys, and the reverse for humans, we put the data from monkeys and humans onto the same scale by calculating a composite performance measure: comparison cost (RT/accuracy), also called "inverse efficiency" (Townsend & Ashby, 1978, 1983). This performance measure is commonly used in the field of visual cognition (Goffaux, Hault, Michel, Vuong, & Rossion, 2005; Jacques & Rossion, 2007; Joassin, Maurage, Campanella, & Bruyer, 2006; Kennett, Eimer, Spence, & Driver, 2001; Murphy & Klein, 1998). The comparison cost was calculated for each stimulus pairing for each subject, in order to test for effects of semantic distance on subjects' responses and compare

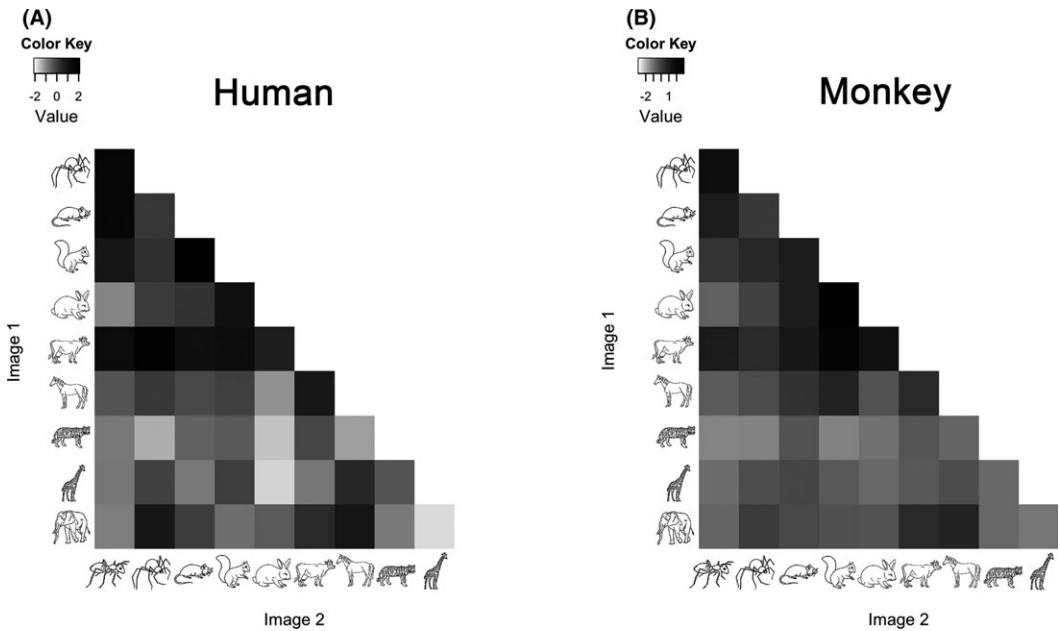


Fig. 2. Representational similarity matrices for performance on animal line drawing stimulus pairs. Shade darkness increases with standardized comparison cost score. (A) Human ($n = 10$) similarity matrix. (B) Monkey ($n = 4$) similarity matrix.

groups. Performance in monkeys was similar regardless of which stimulus of the pair was the target (Spearman's $r = .44$; $p < .01$), so we collapsed data across the 45 unique stimulus pairs. Both monkeys and humans showed a relation between the human-derived semantic distance metrics and their ability to make object judgments (Human: Spearman's $r = .35$, $p < .05$; Monkey: Spearman's $r = .42$, $p < .01$; see Fig. 2 for representational similarity matrices).

Additionally, the pattern of comparison cost across stimulus pairs was highly similar between humans and monkeys (Spearman's $r = .86$, $p < .01$; Fig. 3). We found the same results when we compared monkey and human performance by calculating the correlation between monkey accuracy and human RT (Spearman's $r = -.77$; $p < .01$), and the correlation between the mean of z -scored error rate and RT (Spearman's $r = .86$; $p < .01$; see Section 2).

Finally, we examined whether the similarity between human and monkey performance could be due to the particular properties of the images we used (e.g., one image could be very distinctive from all the others, potentially driving the significant correlation between monkey and human performance). We calculated the marginal total for each image by adding together the comparison costs for each target-distractor pair that included that image. To account for these marginal totals in our measure of performance, we conducted a linear regression predicting comparison cost score from the marginal totals for each image in the pair. The residuals from this multiple regression reflect performance while

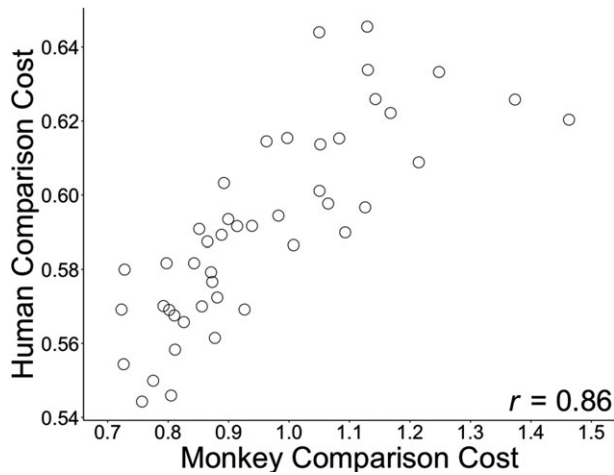


Fig. 3. Human comparison cost scores are plotted against monkey comparison cost scores for each target-distractor animal line drawing pair. Spearman's $r = .86$, $p < .01$.

accounting for the particular properties of the images we used. We obtained residuals for monkey and human performance separately and then correlated them to see if performance was still similar across species. Even when accounting for the marginal totals of the images, monkey, and human performance were still significantly related (Spearman's $r = .72$; $p < .01$). This result indicates that the representational similarity matrices within the category "animals" are similar for monkeys and humans.

We tested whether the relation between semantic distance and performance in monkeys and humans could be explained by dimensions of low-level visual similarity. In order to provide a robust test of the possible contribution of visual dimensions, we tested five measures of visual similarity: Affine Similarity, Shape Context Similarity, Pixel Correlation, Pixel Sum of Squared Differences-inverted (Pixel SSD-inv), and Radon Correlation (see Section 2 for details). Of these five measures, only one (Pixel SSD-inv) showed a significant relation with humans' and monkeys' performance (Pixel SSD-inv: Human: Spearman's $r = .54$, $p < .01$; Monkey: $r = .69$, $p < .01$; all other $ps = .28$ – 1.0 ; alphas corrected for multiple comparisons; see Table S4). We thus conducted a multiple regression analysis to determine the relative contributions of semantic similarity and Pixel SSD-inv to performance. For both humans and monkeys, semantic similarity and Pixel SSD-inv each explained a significant amount of the variance in performance (Human: semantic similarity $sr^2 = .124$, $p < .01$; Pixel SSD-inv $sr^2 = .255$, $p < .01$; Monkey: semantic similarity $sr^2 = .167$, $p < .01$; Pixel SSD-inv $sr^2 = .379$, $p < .01$). That is, even when accounting for Pixel SSD-inv, the relation between semantic similarity and performance remained significant. This result indicates that human and non-human primate representations of objects scale according to the same variables, one of which is human-defined semantic similarity.

In a second experiment, we tested whether the same patterns would be obtained when the stimuli were grayscale photographs (Fig. 1B). In this new stimulus set, the animals were presented as novel photographs. Both groups performed above chance with photographs (Monkey: 84%, $SD = 10\%$; Human: 99%, $SD = 0.5\%$) and with rapid response times (Median RTs: Monkey = 550 ms, $SD = 270$ ms; Human = 550 ms, $SD = 182$ ms). Critically, as was observed with the line drawing stimuli, both monkeys and humans showed a relation between human-derived semantic similarity metrics and comparison cost across stimulus pairings (Human: Spearman's $r = .31$, $p < .05$; Monkey: Spearman's $r = .42$, $p < .01$; see Fig. 4 for representational similarity matrices). Moreover, the degree of comparison cost across stimulus pairs was again highly correlated between monkeys and humans, implicating common scaling of semantic distance among the animal stimuli in humans and monkeys (Spearman's $r = .39$, $p < .01$; see Fig. 5). We again found similar results when we correlated monkey accuracy and human RT (Spearman's $r = -.34$; $p < .05$), the average of z -scored error rate and RT (Spearman's $r = .33$; $p < .05$), and the marginal totals (Spearman's $r = .26$; $p = .08$), comparable to the results from the line drawing task.

As was done for line drawings, we then examined the possible contribution of visual dimensions of similarity to task performance. In addition to the five visual measures examined previously with the animal line drawing stimuli, we also converted the grayscale pixel values to binary values and then used correlations and sum of squared

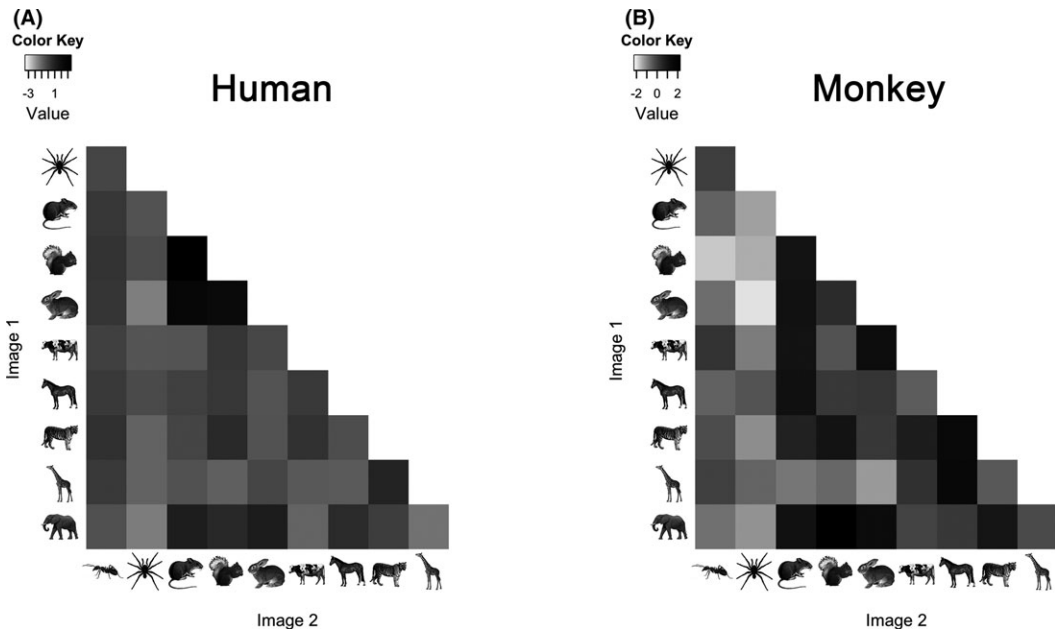


Fig. 4. Representational similarity matrices for performance on animal photograph stimulus pairs. Shade darkness increases with standardized comparison cost score. (A) Human ($n = 10$) similarity matrix. (B) Monkey ($n = 2$) similarity matrix.

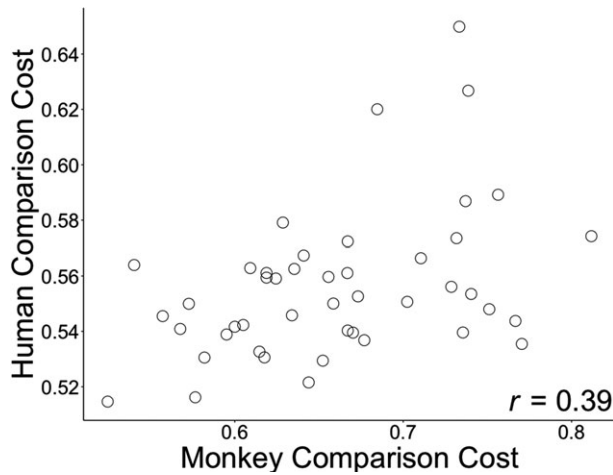


Fig. 5. Human comparison cost scores are plotted against monkey comparison cost scores for each target-distractor animal photograph pair. Spearman's $r = .39$, $p < .01$.

differences to compare the images. Only Binary Pixel SSD-inv was significantly correlated with monkeys' performance after correction for multiple comparisons (Spearman's $r = .42$, $p < .05$; all other $ps = .12$ – 1.0 ; see Table S4). We conducted a multiple regression analysis to determine the relative contributions of semantic similarity and Binary Pixel SSD-inv to monkeys' performance. Only semantic similarity explained a significant amount of the variance in performance (semantic similarity $sr^2 = .086$, $p < .05$; Binary Pixel SSD-inv $sr^2 = .026$, $p = .26$).

In summary, low-level visual similarity is not as good a metric for predicting monkey or human object discrimination as semantic similarity across line drawings and photographs.

As a more integrated analysis of the roles of low-level visual representations and abstract perceptual representations in performance, we tested the behavioral data against the performance of a computational model of visual object processing: HMAX (Riesenhuber & Poggio, 1999b; Serre, Oliva, et al., 2007; Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007; Serre et al., 2005). The HMAX model represents low-level visual (V1) and abstract whole object representations (IT) as separate stages of processing within a hierarchical and feed-forward architecture. Thus, we can test whether monkeys' and humans' comparison cost patterns across the animal stimuli were more similar to the model's estimation of V1 (HMAX-S1C1) or IT (HMAX-C2) representational similarity. To do this, we calculated the similarity between all pairs of images from our stimulus set using the simulated visual processes corresponding to V1 and IT in the HMAX architecture (see Section 2 for details).

Monkey performance on both line drawings and grayscale photos was significantly predicted by the IT outputs (line drawings: Spearman's $r = .54$, $p < .01$; grayscale photos: $r = .38$, $p < .01$), but not V1 outputs (line drawings: $r = .25$, $p = .10$; grayscale photos:

$r = .24, p = .11$). This suggests that monkeys' confusion among animal pairings was driven by abstract, whole object perceptual similarity rather than low-level feature-based similarity.

We then tested whether semantic similarity explained unique variance in discrimination performance when accounting for variance explained by the IT outputs of the HMAX model. To that end, we conducted an ANCOVA over monkeys' and humans' performance on the line drawings and grayscale photographs. We examined the effects of species (monkey or human), stimulus type (line drawing or photograph), semantic similarity, IT-based similarity, and V1-based similarity. The analysis revealed independent effects of semantic similarity and IT-based similarity on performance but no relation between performance and the V1-based similarity (semantic similarity: $F(1, 168) = 18.96, p < .001$; IT model: $F(1, 168) = 7.31, p < .01$; V1 model: $F(1, 168) = 1.99, p = .16$). There was also an effect of species ($F(1, 168) = 265.21, p < .001$) because humans had overall lower comparison costs than monkeys. Follow-up regression analyses showed that semantic similarity still explained variance in performance, after accounting for the contribution of the IT-based similarity within each species (Monkeys: semantic similarity: $sr^2 = .05, p < .01$; IT model: $sr^2 = .02, p < .05$; Humans: semantic similarity: $sr^2 = .07, p < .01$; IT model: $sr^2 = .09, p < .01$).

Together, these results confirm that high-level similarities among objects significantly influence the judgments of monkeys and humans, above and beyond the similarity of visual representations.

3.2. Match-to-sample task with novel object stimuli

Overall, the data indicate that human-defined semantic distance is a strong and independent predictor of monkeys' discriminations of animal images beyond established measures of visual similarity: Affine Similarity, Shape Context Similarity, Pixel Correlation, Pixel Sum of Squared Differences-inverted (Pixel SSD-inv), Correlation of Radon Transformed Images (Radon Correlation), Otsu's visual similarity, and V1 (HMAX-S1C1) and IT (HMAX-C2) components. An important positive control would thus be to demonstrate that these indices of visual similarity are able to capture monkey performance when the stimuli do not invoke abstract object representations. To that end, we tested monkeys on a match-to-sample task, using novel geometric stimuli. The stimuli were grayscale renderings of artificial objects (details in Section 2—see Fig. 1C). Monkeys performed the task well above chance ($p < .01, 76\%$ accuracy, $SD = 1\%$), and with rapid response times (Median RT = 617 ms, $SD = 330$ ms). Comparison cost was calculated, using the same method as for the animal line drawing and photograph tasks. Comparison costs across the object pairings were highly correlated with nearly all of the visual similarity measures: Pixel Correlation (Spearman's $r = .58, p < .01$), Pixel SSD-inv (Spearman's $r = .58, p < .01$), Radon Correlation (Spearman's $r = .57, p < .01$), Binary Pixel Correlation (Spearman's $r = .60, p < .01$), and Binary Pixel SSD-inv (Spearman's $r = .57, p < .01$) (corrected for multiple comparisons; see Table S4 for complete correlation results). As with the previous tasks, we tested performance against the V1 and IT levels of the

HMAX model of visual object processing. We used a multiple regression to determine the unique contribution of the V1 and IT models to monkeys' performance. The V1 model explained unique variance in performance ($sr^2 = .17, p < .01$), but the IT model did not ($sr^2 = .02, p = .31$). These results show that when monkeys make judgments over novel object stimuli that have little semantic value, visual similarity largely predicts their performance. This result stands in contrast to the monkeys' performance with animal stimuli, which showed strong relations to semantic similarity and IT-based similarity, but only isolated and weak relations with lower level visual similarity. These data serve the important function of demonstrating that the measures of visual similarity we used have the necessary sensitivity to explain monkeys' performance, when their performance is principally driven by an analysis of low-level visual information.

3.3. *Picture–word agreement task with animal stimuli*

In both the line drawing and grayscale image versions of the matching task, we observed that humans' and monkeys' discrimination abilities were correlated across stimulus pairs. This suggests similar scaling of objects from the category "animals" in humans and monkeys. In order to confirm that the object representations that humans used in this task are representative of the abstract representations that humans invoke in verbal tasks, we tested humans on a picture–word agreement task with the same stimuli tested in the first two experiments. A new group of human subjects was given a task in which the animal line drawings (Fig. 1A) were paired with written animal names. Subjects reported whether the picture and word represented the same or a different animal. Human subjects' comparison cost scores were calculated from the non-match trials (overall error rate = 3%). Importantly, human performance across these picture–word animal pairs was related to monkeys' performance across the same animal pairs from the line drawing task as measured by comparison cost (Spearman's $r = .28, p = .06$; Pearson's $r = .46, p < .01$) and the correlation between monkey accuracy and human RT (Spearman's $r = -.29, p < .05$). This indicates that even when human subjects are judging objects using abstract verbal stimuli, their representations are scaled similarly to those of non-human primates.

4. Discussion

There has been substantial research indicating homologies between humans and monkeys in the neural systems underlying visual processing. However, there has been limited research linking those known neurobiological homologies to homologies in cognition and behavior. It is important to show representational similarities behaviorally because only the animals' behavior shows how they use object information to make judgments. Here, we show across three datasets (line drawings, photographs, and words) that within-category semantic scaling for the category "animals" is highly similar between monkeys and humans.

Human-derived measures of the semantic similarity and whole-object similarity (HMAX IT-based) predicted monkeys' and humans' discrimination of animals across two experiments: one with line drawings and another with photographs. We further showed that the relation between human-derived measures of semantic similarity and discrimination performance could not be explained purely by simple visual similarity among the stimuli (Affine Similarity, Pixel Similarity, Radon transform, V1-based, etc.). In contrast, simple visual similarity did predict monkeys' performance when the task was performed over artificial geometric objects. These data provide evidence of common influences of whole object and semantic similarity in the "animal" representations of humans and non-human primates. This is surprising because human concepts of "animals" draw on encyclopedic and formal knowledge made possible by language, and which is unavailable to non-humans. Moreover, the monkeys in this study are captive and have no experience with the animals pictured in the stimuli. Our findings indicate that information from the animal images provides humans and animals with a common representational space accounted for by whole object similarity and non-verbal semantic knowledge. Research on object representation in IT cortex, the most abstract stage in the object recognition pathway, supports a representational space based on whole object similarity. For example, some IT neurons respond selectively and invariantly to complex shapes or even hands or faces (Desimone, Albright, Gross, & Bruce, 1984; Gross, Rocha-Miranda, & Bender, 1972). Since the monkeys have no direct or encyclopedic knowledge of the animals in the stimuli, any semantic knowledge would have to be extrapolated or inferred from prior experience with other objects and the target object's perceptual features like shape and composition.

Although we examined the relation between task performance and many measures of low-level visual similarity, additional low-level measures (e.g., a silhouette model) should be examined in future work. Additionally, there is disagreement about the extent to which the HMAX model used here represents cells in IT cortex. Future work should include visual similarity measures based on alternative representations of IT cortex, such as deep learning models like CAFFE (Khaligh-Razavi & Kriegeskorte, 2014). Regardless of how well the HMAX model represents IT, the model's output represents images as sets of features, in a manner tolerant to shifts and size differences. Our findings that monkey and human performances with animal stimuli were related to similarity based on this more abstract representation of the stimuli, rather than measures of low-level visual similarity, support our conclusion that monkeys and humans represent these animal stimuli at a level of perceptual or categorical abstraction that goes beyond low-level visual properties.

It is perhaps surprising that the measures of low-level visual similarity among the animal stimuli were, overall, weak predictors of monkey performance. Important in this context is the positive control experiment demonstrating the efficacy of those measures for predicting monkey performance when the monkeys were making discriminations over geometric (i.e., meaningless) stimuli. Those data indicate that the types of representations that drove performance when the stimuli were images of animals (high-level perceptual) were of a fundamentally different nature than the representations that drove performance when the stimuli were novel geometric objects (low-level visual). The implication is that the animal stimuli are represented at a more "meaningful," abstract perceptual or

categorical level compared to the arbitrary geometric stimuli. It is unclear what caused the monkeys to analyze these stimuli differently. The novel geometric stimuli are visually quite different from the animal stimuli, so the monkeys' different treatment of those stimuli could be related to visual feature differences between the stimulus categories. Different levels of analysis for the animal stimuli versus geometric shapes could be triggered by either the greater semantic content in the animal stimuli or the salient, simple visual properties of the geometric stimuli.

Although our data establish a fundamental similarity in the object representations of humans and non-human primates, the precise nature of the abstract representations that monkeys accessed during the task remains to be determined. One possibility is that monkeys and humans have similar biases for weighting core object features to determine their subjective similarity relations. Resolution of this issue is important not only for determining the nature of abstract representations in primates but also for understanding the informal, non-experiential, and non-verbal basis of human semantic structure. The fact that naïve monkeys' judgments of animal stimuli were highly correlated with human judgments of the same stimuli (in both verbal and non-verbal tasks) raises the possibility that humans recruit a more primitive form of object representation than would be expected from the formal and language-based knowledge that is available to them. This possibility is supported by research on geometric shape processing showing that humans process shape information at different depths depending on whether the stimuli are presented verbally or pictorially (Edwards, Boyer, Bell, & Sturz, 2016; Sturz, Edwards, & Boyer, 2014). The human ability to access a primitive form of object representation that is similar to the representations of non-human primates would be predicted by evolutionary constraints on human object perception (Dehaene & Cohen, 2007; Geary, 2005; Mahon & Caramazza, 2011; New et al., 2007; Orians & Heerwagen, 1992; Santos & Caramazza, 2002; Spelke & Kinzler, 2007). The idea is that humans retain a primitive system of object representation, common to non-human primates, that is exapted by modern humans to aid in the acquisition of evolutionarily more recent formal and cultural knowledge.

Together, our results show that high-level "semantic" similarities among objects significantly influence the judgments of monkeys and humans, above and beyond the similarity of low-level visual features. The shared representations of monkeys and humans do not depend on direct knowledge of or experience with the objects since the monkeys had none. Instead, the common metric of representational scaling we have reported between humans and monkeys implicates a key role for evolutionary constraints in the organization of human object knowledge.

Acknowledgments

We thank Stephen Ferrigno for assistance conducting the match-to-sample experiments, Alena Stasenko for conducting the picture–word experiment, Robert Jacobs and Matthew Overlan for the novel object stimuli, and Steven Piantadosi for consulting on statistical analyses. This material is based upon work supported by the National Science Foundation

Graduate Research Fellowship under grant no. DGE-1419118 (to SEK), the James S. McDonnell Foundation, and the Alfred P. Sloan Foundation (to JFC). BZM was supported, in part, by National Science Foundation grant 1349042 and NIH grant R01 NS089609.

References

- Aydelott, J., Baer-Henney, D., Trzaskowski, M., Leech, R., & Dick, F. (2012). Sentence comprehension in competing speech: Dichotic sentence-word priming reveals hemispheric differences in auditory semantic processing. *Language and Cognitive Processes*, 27(7–8), 1108–1144. doi:10.1080/01690965.2011.589735
- Belongie, S., Belongie, S., Malik, J., Malik, J., Puzicha, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(24), 509–522. doi:10.1.1.18.8852->
- Bovet, D., & Vauclair, J. (2001). Judgment of conceptual identity in monkeys. *Psychonomic Bulletin & Review*, 8(3), 470–475. doi:10.3758/BF03196181
- Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, 282(5389), 746–749. doi:10.1126/science.282.5389.746
- Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, 17(5), 401–406. doi:10.1111/j.1467-9280.2006.01719.x
- Caramazza, A., Hersh, H., & Torgerson, W. S. (1976). Subjective structures and operations in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 15(1), 103–117. doi:10.1016/S0022-5371(76)90011-6
- Carey, S. (1982). Semantic development: The state of the art. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 347–389). Cambridge, UK: Cambridge University Press.
- Cerella, J. (1979). Visual classes and natural categories in the pigeon. *Journal of Experimental Psychology: Human Perception and Performance*, 5(1), 68–77. doi:10.1037/0096-1523.5.1.68
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428.
- Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132(2), 163–201. doi:10.1037/0096-3445.132.2.163
- Crowley, J. L. (1997). Vision for man-machine interaction. *Robotics and Autonomous Systems*, 19(3–4), 347–358. doi:10.1016/S0921-8890(96)00061-9->
- Dehaene, L., & Cohen, S. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384–398. doi:10.1016/j.neuron.2007.10.004
- Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *The Journal of Neuroscience*, 4(8), 2051–2062.
- Edwards, J. E., Boyer, T. W., Bell, Z. K., & Sturz, B. R. (2016). Isolated processing of geometric shapes and their corresponding shape words: Evidence from a delayed match-to-sample task. *Journal of Experimental Psychology: Human Perception and Performance*, 42(8), 1088–1103.
- Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *NeuroReport*, 9(2), 303–308. doi:10.1097/00001756-199801260-00023
- Fize, D., Cauchoix, M., & Fabre-Thorpe, M. (2011). Humans and monkeys share visual representations. *Proceedings of the National Academy of Sciences*, 108(18), 7635–7640. doi:10.1073/pnas.1016213108
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44(1), 43–74.
- Geary, D. C. (2005). *The origin of mind: Evolution of brain, cognition, and general intelligence*. Washington, DC: American Psychological Association.

- Gil-da-Costa, R., Braun, A., Lopes, M., Hauser, M. D., Carson, R. E., Herscovitch, P., & Martin, A. (2004). Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(50), 17516–17521. doi:10.1073/pnas.0408077101
- Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*(1), 77–86. doi:10.1068/p5370
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research* *41*, 1409–1422, doi:10.1016/S0042-6989(01)00073-6
- Gross, C. G., Rocha-Miranda, C. E. D., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, *35*(1), 96–111.
- Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, *146*(October), 549–551. doi:10.1126/science.146.3643.549
- Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*(4), 285–302. doi:10.1037//0097-7403.2.4.285
- Humphreys, G. W., Price, J., & Riddoch, M. J. (1999). From objects to names : A cognitive neuroscience approach. *Psychological Research*, *62*(2–3), 118–130.
- Jacques, C., & Rossion, B. (2007). Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *NeuroImage*, *36*(3), 863–876. doi:10.1016/j.neuroimage.2007.04.016
- Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, doi:10.1111/j.1365-2929.2004.02012.x
- Joassin, F., Maurage, P., Campanella, S., & Bruyer, R. (2006). Is associative priming a valid method to differentiate the serial and parallel models of face identification? *Visual Cognition*, *14*(2), 199–216. doi:10.1080/13506280544000264
- Kanade, T., & Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*(9), 920–932. doi:10.1109/34.310690
- Keil, F. (1981). Children’s thinking: What never develops? *Cognition*, *10*, 159–166. doi:10.1016/0010-0277(81)90040-8
- Kennett, S., Eimer, M., Spence, C., & Driver, J. (2001). Tactile-visual links in exogenous spatial attention under different postures: Convergent evidence from psychophysics and ERPs. *Journal of Cognitive Neuroscience*, *13*(0898-929X), 462–478. doi:10.1162/08989290152001899
- Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), e1003915. doi:10.1371/journal.pcbi.1003915
- Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230. doi:10.1167/11.11.923
- Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. doi:10.1016/j.tics.2012.10.011
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 4. doi:10.3389/neuro.06.004.2008
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K. . . . , & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. doi:10.1016/j.neuron.2008.10.043
- Logothetis, N. K., Guggenberger, H., Peled, S., & Pauls, J. (1999). Functional imaging of the monkey brain. *Nature Neuroscience*, *2*(6), 555–562. doi:10.1038/9210

- Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? The distributed domain-specific hypothesis. *Trends in Cognitive Sciences*, *15*(3), 97–103. doi:10.1016/j.tics.2011.01.004.
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(3), 503–535. doi:10.1037/0278-7393.33.3.503
- Maki, W. S., & Buchanan, E. M. (2008). Latent structure in measures of associative, semantic, and thematic knowledge. *Psychonomic Bulletin & Review*, *15*(3), 598–603. doi:10.3758/PBR.15.3.598
- Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition: The Minnesota symposia on child psychology* (Vol. 25 pp. 59–101). Hillsdale, NJ: Erlbaum.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. doi:10.1146/annurev.psych.57.102904.190143
- Martin, A. (2009). Circuits in the mind: The neural foundations for object concepts. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1031–1045). Cambridge, MA: MIT Press.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. doi:10.3758/BRM.40.1.183
- Mirman, D. (2011). Effects of near and distant semantic neighbors on word production. *Cognitive, Affective & Behavioral Neuroscience*, *11*(1), 32–43. doi:10.3758/s13415-010-0009-7
- Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 65–79. doi:10.1037/0278-7393.34.1.65
- Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition*, *37*(7), 1026–1039. doi:10.3758/MC.37.7.1026
- Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*, 128.
- Murphy, F. C., & Klein, R. M. (1998). The effects of nicotine on spatial and non-spatial expectancies in a covert orienting task. *Neuropsychologia*, *36*(11), 1103–1114. doi:10.1016/S0028-3932(98)00012-8
- Neiworth, J. J., & Wright, a. a. (1994). Monkeys (*Macaca mulatta*) learn category matching in a nonidentical same-different task. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*(4), 429–435. doi:10.1037/0097-7403.20.4.429
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(42), 16598–16603. doi:10.1073/pnas.0703913104
- Orians, G. H., & Heerwagen, J. H. (1992). Evolved responses to landscapes. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *Adapted mind: Evolutionary psychology and the generation of culture* (pp. 555–579). New York: Oxford University Press. doi:citeulike-article-id:7333260
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* *9*(1), 62–66.
- Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 109–178.
- Phillips, L. H., & Rabbit, P. M. A. (1995). Impulsivity and speed-accuracy strategies in intelligence test performance. *Intelligence*, *21*(1), 13–29. doi:10.1016/0160-2896(95)90036-5
- Phillips, W., & Santos, L. R. (2007). Evidence for kind representations in the absence of language: Experiments with rhesus monkeys (*Macaca mulatta*). *Cognition*, *102*(3), 455–463. doi:10.1016/j.cognition.2006.01.009

- Riesenhuber, M., & Poggio, T. (1999a). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019–1025. doi:10.1038/14819
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12(1), 1–20. doi:10.1016/S0022-5371(73)80056-8
- Sands, S. F., Lincoln, C. E., & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *Journal of Experimental Psychology: General*, 111, 369–389. doi:10.1037/0096-3445.111.4.369
- Santos, L., & Caramazza, A. (2002). The domain-specific hypothesis. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category specificity in brain and mind* (pp. 1–24). New York: Psychology Press.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 104(15), 6424–6429. doi:10.1073/pnas.0700622104
- Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), 411–426. doi:10.1109/TPAMI.2007.56
- Serre, T., Wolf, L., & Poggio, T. (2005). Object recognition with features inspired by visual cortex. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2, 994–1000.
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, 133(2), 166–188.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, 6(2), 174–215. doi:10.1037/0278-7393.6.2.174
- Spelke, E. S., Bernier, E. P., & Skerry, A. E. (2013). Core social cognition. In M. Banaji & S. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 11–16). New York: Oxford University Press.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10, 89–96. doi:10.1111/j.1467-7687.2007.00569.x
- Sturz, B. R., Edwards, J. E., & Boyer, T. W. (2014). Asymmetrical interference effects between two-dimensional geometric shapes and their corresponding shape words. *PLoS ONE*, 9(3), e92740. doi:10.1371/journal.pone.0092740
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation. *Current Anthropology*, 53(6), 673–692. doi:10.1086/668207
- Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. *Cognitive Theory*, 3, 200–239.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, MA: Cambridge University Press.
- Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6(9), 989–995. doi:10.1038/nn1111
- Van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, 13(1), 1–10. doi:10.1016/0896-6273(94)90455-3
- Vauclair, J., & Fagot, J. (1996). Categorization of alphanumeric characters by Guinea baboons: Within- and between-class stimulus discrimination. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition*, 15(5), 449–462. Retrieved from <Go to ISI>://WOS:A1996VT85200001.
- Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48(4), 422–488. doi:10.1016/j.cogpsych.2003.09.001
- Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 1: Behavioural study. *European Journal of Neuroscience*, 11(4), 1223–1238. doi:10.1046/j.1460-9568.1999.00530.x

- Wade, A. R., & Tyler, C. W. (2005). Human lateral occipital cortex contains a non-retinotopic map of visual space. Proceedings of the Annual Meeting of the Organization for Human Brain Mapping, Toronto, Canada.
- Wandell, B. A. (1999). Computational neuroimaging of human visual cortex. *Annual Review of Neuroscience*, 22(February), 145–173. doi:10.1146/annurev.neuro.22.1.145
- Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge. *Brain*, 110(5), 1273–1296.
- Wasserman, E. a., Kiedinger, R. E., & Bhatt, R. S. (1988). Conceptual behavior in pigeons: Categories, subcategories, and pseudocategories. *Journal of Experimental Psychology: Animal Behavior Processes*, 14(3), 235–246. doi:10.1037/0097-7403.14.3.235
- Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, 63(2), 165–174. doi:10.1901/jeab.1995.63-165
- Wei, T., Liang, X., He, Y., Zang, Y., Han, Z., Caramazza, A., & Bi, Y. (2012). Predicting conceptual processing capacity from spontaneous neuronal activity of the left middle temporal gyrus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(2), 481–489. doi:10.1523/JNEUROSCI.1953-11.2012

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Pairwise Spearman correlations between the visual similarity measures for the animal line drawing stimuli.

Table S2. Pairwise Spearman correlations between the visual similarity measures for the animal photo stimuli.

Table S3. Pairwise Spearman correlations between the visual similarity measures for the novel geometric object stimuli.

Table S4. Spearman correlations between comparison cost and low-level visual similarity measures for monkeys and humans in each task. *p*-values (in parentheses) are corrected using the Holm–Bonferroni method, within each task and subject group.