

# Similarity and structured representation in human and nonhuman apes

Carl J. Hodgetts<sup>1,2\*</sup>, James O. E. Close<sup>3,4</sup>, & Ulrike Hahn<sup>5</sup>

1. Department of Psychology, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK

2. Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Maindy Road, Cardiff, CF24 4HQ, UK

3. Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103, Leipzig, Germany

4. School of Psychology and Sport Science, Anglia Ruskin University, East Road, Cambridge CB1 1PT, UK

5. Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK

\* For correspondence: [carl.hodgetts@rhul.ac.uk](mailto:carl.hodgetts@rhul.ac.uk)

---

## Abstract

How we judge the similarity between objects in the world is connected ultimately to how we represent those objects. It has been argued extensively that object representations in humans are 'structured' in nature, meaning that both individual features and the relations between them can influence similarity. In contrast, popular models within comparative psychology assume that nonhuman species appreciate only surface-level, featural similarities. By applying psychological models of structural and featural similarity (from conjunctive feature models to Tversky's contrast model) to visual similarity judgements from adult humans, chimpanzees, and gorillas, we demonstrate a cross-species sensitivity to complex structural information, particularly for stimuli that combine colour and shape. These results shed new light on the representational complexity of nonhuman apes, and the fundamental limits of featural coding in explaining object representation and similarity, which emerge strikingly across both human and nonhuman species.

**Keywords:** *similarity, object representation, features, cognitive evolution, perceptual discrimination*

---

## 1 Introduction

Across the animal kingdom exists the capacity to extend familiar behaviours to novel but similar situations and objects. This makes similarity a fundamental concept within models of human and nonhuman cognition (Pearce, 1994; Rescorla & Wagner, 1972). Indeed, it has been shown to influence learning (Ross, 1984), memory (Simons et al., 2005), generalization (Osherson et al., 1990), categorization (Nosofsky, 1984), and even social behaviour (White, 2008). Similarity, however, is fundamentally in the eye of the perceiver (Goldstone, 1994b; Hahn & Chater, 1997; Medin et al., 1993). It is not a property of physical objects themselves, but rather a property of how an animal represents those objects. Specifically, similarity is a function of those aspects of an object that are encoded, and the importance assigned to them. Theories of similarity, therefore, have close connections with theories of representation: how real-world objects are internally represented affects perceived similarity, and perceived similarity, in turn, provides insight into mental representation (Edelman, 1998; Hahn, 2014).

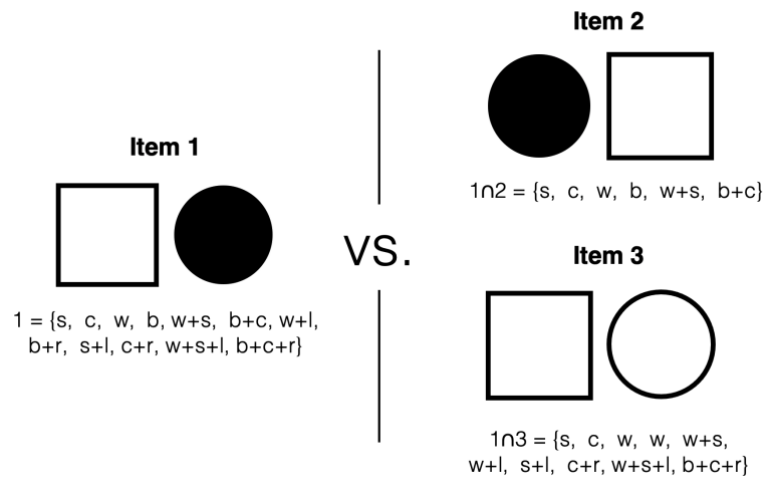
41 A major debate within human psychology concerns the role of *relations* in similarity: a table,  
42 for example, is not just a collection of features (tabletop, legs, colour etc.) but these features  
43 arranged in a particular way. Human visual representation, therefore, is argued to involve so-called  
44 'structured representations', that is, representations that involve both features and the relations  
45 between them (Biederman, 1987; Hafri & Firestone, 2021; Hahn et al., 2003; Markman & Gentner,  
46 1993). Indeed, it has been shown that when two objects share a perceptual feature (e.g., 'red'), it  
47 contributes more to human similarity judgments when it appears in corresponding positions of a  
48 relational structure – referred to as a 'match-in-place', or MIP (Goldstone, 1994a). For instance, if  
49 presented with an image of two people wearing coloured hats and shirts, the pair will be perceived  
50 to be more similar if both hats are red and both shirts blue, than if the hat and shirt colours are  
51 swapped for one person in the pair, despite the overall feature set remaining unchanged.

52 Attempts to formally measure similarity, both within cognitive psychology and machine  
53 learning, have thus started to move toward ways of calculating similarity over structured  
54 representations (Hahn et al., 2003; Markman & Gentner, 1993). However, many of the most popular  
55 models, particularly in the context of animal learning (Pearce, 1994; Rescorla & Wagner, 1972), still  
56 treat stimulus representation as a matter of decomposing stimuli into individual, task-relevant  
57 features, whether in a feature vector (Tversky, 1977), or a spatial representation that represents  
58 items as points in a multi-dimensional space (Shepard, 1957). As we will outline below, the only way  
59 relational structure is encoded in these models is by treating relations themselves as 'features', in  
60 particular, as 'conjunctive features' (i.e., 'red' and 'hat' becomes 'red+hat').

61 While such 'featural' representations appear too simplistic for humans – at least in some  
62 contexts (Goldstone & Medin, 1994; Hahn et al., 2003; Markman & Gentner, 1993) – the widespread  
63 use of featural coding schemes in the animal learning literature would seem to reflect an implicit (or  
64 even explicit) assumption that these may be adequate for some, or even *all*, nonhuman species.  
65 This assumption might have been fuelled further by the finding that nonhuman species have  
66 difficulty recognising purely relational similarities (Blough, 2001; George et al., 2001; Haun & Call,  
67 2009). The fact that nonhuman primates do not seem to be able to deal with such purely relational  
68 similarities (e.g., "equal/unequal", or "sameness") is of course distinct from the question of whether  
69 relational information impacts similarity judgements more generally. Needless to say, if nonhuman  
70 primates do not include relational information in their object representations, then such  
71 information cannot impact their similarity judgements. The fact that they might, however, as has  
72 been argued in some studies (Hopkins & Washburn, 2002; Huber & Lenz, 1993; Kirkpatrick-Steger  
73 et al., 1998; Zentall et al., 2014), does not prejudice if and how this relational information impacts  
74 perceived similarity. Just showing sensitivity to relational structure does not tell us much (if  
75 anything) about the similarity *gradients* that stem from matches or mismatches. To understand this,

76 structural models of similarity from the human literature must be tested systematically on animal  
 77 behaviour. Given the close connection between similarity and theories of representation outlined  
 78 above, such an examination should elucidate not only similarity, in particular the limitations of  
 79 feature-based models of similarity, but may also provide valuable insight into the mental object  
 80 representations of nonhuman species.

81  
 82



83

84 **Figure 1.** A simple set of three stimuli (*items 1, 2 and 3*) for comparison. Below *item 1* is a featural description  
 85 in terms of the basic features ( $s$  = 'square',  $c$  = 'circle',  $w$  = 'white', and  $b$  = 'black'), the feature conjunctions ( $w$   
 86 +  $s$  = 'white+square',  $b$  +  $c$  = 'black+circle', etc.) and the features/conjunctions including relative spatial  
 87 location ( $w+l$  = 'white+left',  $w+s+l$  = 'white+square+left', etc.). To the right are the features that *items 2 and 3*  
 88 share with *item 1*. Note also that the comparison  $1\cap 3$  contains two matches in terms of colour: the colour  
 89 match between the two squares and the colour match between square in the *item 1* and circle in *item 3*. This  
 90 latter match feels like it should 'count for less', in line with prior work (Gentner & Markman, 1997), but on a  
 91 conjunctive scheme is dealt with by the conjunctions 'colour+shape' and 'colour+shape+relative location'.  
 92 Even excluding the multiple match, however, *item 2* remains less similar to *1* than *3*.

93

94

### 95 1.1 The limits of featural coding

96 To address this question, we examined the nature of perceptual similarity judgements in three great  
 97 ape species: humans, chimpanzees, and gorillas. Our main question was whether perceived  
 98 similarity in chimpanzees and gorillas are, like humans, sensitive to 'structure'. To this end, we  
 99 devised a simple stimulus set that allowed us to adjudicate between featural and structural models  
 100 of similarity. Before outlining these stimuli and the model predictions in detail, we will first describe  
 101 some of the fundamental issues facing featural models of similarity. This is best done with reference  
 102 to a particular set of items, so we will use example items from our actual stimulus set. Each item in  
 103 our stimulus set involves a pair of geometric figures, such as those shown in Figure 1. Broadly,  
 104 featural models of similarity treat each psychologically relevant aspect of an object as a single

105 component, and the entire object is represented by all relevant components - whether these are  
106 represented as a feature set, a feature vector, or a point in a multi-dimensional feature space  
107 (Shepard, 1980; Tversky, 1977). To illustrate: the left stimulus in Figure 1 (Item 1) might be  
108 represented by the feature set {square, circle, white, black}.

109 However, it is not just the simple attributes such as colour or shape that are potentially  
110 relevant to similarity but also relations between these attributes: a particular colour and shape are  
111 bound together in the same component object (e.g., the square), and that object is arranged in a  
112 particular way relative to the second object (the circle). Featural representations capture such  
113 relational information only by turning the relation itself into a 'feature'. This type of conjunctive  
114 coding introduces a feature 'black+circle' to capture the fact that it is the circle not the square that is  
115 black, thus expanding the feature set to {white, black, square, circle, white+square, black+circle}.

116 Such an approach runs into trouble because it leads to a proliferation of 'features': all the  
117 basic elements *and* all their possible combinations must be retained to avoid the hyper-specificity  
118 combinations would otherwise bring. To illustrate, if one were to only consider 'white+square+left'  
119 as a single compound feature, then *item 1* and *item 2* no longer share any features and would thus  
120 appear maximally dissimilar. In other words, one needs to retain both the component features  
121 ('white') and the conjunctions ('white+square') to account for commonalities that appear across  
122 different positions and/or objects. This requirement not only leads to a combinatorial explosion  
123 (with the number of conjunctions determined by the binomial coefficient ' $N$  choose  $k$ ', which for 10  
124 individual features will add a further 45 separate feature pairs, and at  $N = 15$  a further 105) but all of  
125 these features potentially influence the similarity comparison and counting them in assessing  
126 'commonality' itself can lead to counter-intuitive distortions. For example, such proliferations make  
127 it the case that *items 1* and *2* have considerably *fewer* features in common when compared to *items*  
128 *1* and *3*.

129 Representation schemes that allow one to represent relations are known as structured  
130 representations (Gentner, 1983; for more recent literature, see e.g., (Doumas et al., 2008; Doumas  
131 & Martin, 2018; Poldrack, 2020; Shepherd, 2018). The most popular examples of such schemes are  
132 graph structures or multi-place predicates as found in first-order logic, for example, a 2-place  
133 predicate such as TO-THE-LEFT-OF( $x,y$ ). Turning relations into features means that a binary  
134 relation such as TO-THE-LEFT-OF( $x,y$ ) effectively becomes a feature such as TO-THE-LEFT-OF-  
135 X( $y$ ). The crucial difference between these two schemes is that only the former, relational  
136 representation separates out the relation and both its arguments in such a way that they can be  
137 accessed simultaneously and thus independently factored into the similarity comparison. TO-THE-  
138 LEFT-OF( $x,y$ ) might, for example, provide a relational match to TO-THE-LEFT-OF( $q,r$ ) thus  
139 providing a purely relational commonality across multiple object pairs such as those in Figure 1. For

140 the feature-based version of the same state of affairs, one is left simply with two distinct properties  
141 TO-THE-LEFT-OF-X( ) and TO-THE-LEFT-OF-Y( ). The same is ultimately true of the kinds of  
142 conjunctive coding schemes that are popular in associative and connectionist models (Blumberg &  
143 Sokoloff, 2001; Dickinson, 2012; Gluck & Bower, 1988).

144 It is for this reason that many theories concerned with the representation of real-world  
145 objects or events - whether these are faces, scenes, sentences, or extended narratives - assume that  
146 these cannot be represented on purely featural schemes (Biederman, 1987; Hahn et al., 2003;  
147 Markman & Gentner, 1993). Instead, they seem to require structured representations: complex  
148 representations of objects, their parts and properties, and – crucially - the interrelationships  
149 between them that cannot be boiled down to either lists of features or points in space.

150 That said, it is extremely difficult empirically to distinguish between representation  
151 schemes and with them different approaches to measuring similarity. As the examples of Figure 1  
152 illustrate, similarity depends on representation. If only basic, elemental features ('square', 'circle',  
153 etc.) are considered, but no conjunctions, for example, then *items 2 and 3* are equally similar to *item*  
154 *1*. Without independent specification of the representation of a pair of items, any degree of  
155 similarity between them can likely be generated simply by 'tweaking' the representation. However,  
156 our understanding of human cognition (let alone nonhuman cognition) is simply not advanced  
157 enough to provide a sufficiently detailed, and independent, specification of those representations.

158 Without independent constraint, however, even very general contrasts, such as evidence  
159 simply *for structure* over purely featural representations becomes incredibly difficult, because  
160 representational flexibility allows different types of account to mimic each other's predictions. Even  
161 a classic finding indicating the importance of structure in similarity judgments, such as the larger  
162 effect on similarity of features that appear within corresponding positions of a relational structure  
163 (i.e., matches-in-place), is subject to this. The example of the hats and sweaters from above can be  
164 captured readily through the assumption of conjunctive features ('red+hat', 'blue+sweater') in  
165 addition to the basic features 'red', 'blue', 'sweater' and 'hat'. Matches-in-place simply give rise to  
166 both a match in terms of elements, and to a match in terms of conjunctive features. Without  
167 simultaneously providing evidence *against* conjunctive coding as a sufficient, alternative  
168 explanation (which those studies do not provide), demonstrating the effect of MIPs may provide  
169 only rather weak evidence for structural representations.

170 Though difficult, we seek to show that such evidence against conjunctive coding is possible.  
171 The strategy for doing so is already hinted at in Figure 1. Specifically, the contrast between featural  
172 and structural representation can be pursued successfully over a suitable *set* of items. While  
173 'featural' reconstructions may seem *locally plausible* when considering just one or two comparisons,  
174 they can be shown to be *globally implausible* over an entire set of items. The crucial 'trick' here is to

175 take whatever conjunctive features 'do the work' in one comparison and then identify another  
 176 contrast that can be added to the set of similarity comparisons under consideration where those  
 177 conjunctive features lead to difficulties, generating implausible predictions of similarity. The  
 178 stimulus set on which the empirical work in this paper is presented was designed to do just that.








179

## 180 1.2 Devising a stimulus set for contrasting featural and structural models

181 Our stimulus materials are pairs of simple shapes and are based on a domain that has been  
 182 successfully applied across species to study similarity (Hodgetts et al., 2009; Hodgetts & Hahn,  
 183 2012; Larkey & Markman, 2005), feature binding (Cherries et al., 2006) and analogical reasoning  
 184 (Fagot & Thompson, 2011; Vonk, 2003). We constructed from this domain a set of items that would  
 185 distinguish structural accounts from a variety of possible feature models. These stimuli are shown in  
 186 Table 1. As can be seen, each stimulus comprises two coloured geometric figures which are always  
 187 compared to the same reference (or target) stimulus. The fact that these items are a composite of  
 188 two shapes makes it possible to readily manipulate featural and relational attributes of the stimulus.  
 189 Our set involves seven such comparisons in total (labelled A to G), which vary systematically along  
 190 two dimensions: shape and colour (Task 1) and shape and inner line orientation (Task 2; see  
 191 Methods and Figure 2A for more detail). For clarity, we will refer to only to the Task 1 items below.  
 192 Table 1 summarises the model predictions for each comparison. We next outline these models.

193

194

Comparison	FEAT	C-FEAT	CS-FEAT	S-FEAT	MIP	RD
A 	4*	6*	12	4	4	0
B 	4*	6*	6*	0*	4*	1
C 	4*	4	6	2	2	1
D 	4*	5	9*	3	3*	2
E 	4*	6	9*	2	2*	4
F 	3	4	4	0*	3	3
G 	0	0	0	0*	0	8

195

196 **Table 1.** Predicted similarities for our stimulus comparisons across a range of featural and structural models.  
 197 The stimulus comparisons A-G are listed in the first column and are further described in Figure 2. The  
 198 remaining columns refer to the different models tested. The first four models listed are 'featural' models: a  
 199 basic feature model (FEAT), a feature model that codes feature conjunctions (C-FEAT), a conjunctive model  
 200 that also codes relative spatial location (CS-FEAT), and a feature model that matches features only if they  
 201 appear in corresponding spatial positions (S-FEAT). Additional information about how these model  
 202 predictions were derived can be found in the Supplementary Methods. Comparisons marked with an asterisk  
 203 are discussed in the main text.

204

205 The first model considers only basic feature matches (see **FEAT**, Table 1). Feature matches are  
206 counted on each stimulus dimension separately (e.g., colour and shape) and can be matched  
207 multiple times (e.g., the feature 'blue' in the target item forms a match with both blue features in  
208 comparison *E*). Critically, this basic feature model cannot distinguish between comparisons that  
209 share the same features but in different spatial arrangements, that is, comparisons *A* to *C*. Given  
210 that these three comparisons seem perceptually distinct, at least for humans, a purely featural  
211 representation of this kind seems insufficient. At the very least, colour and shape need to be bound  
212 together into coherent objects, reflecting the fact that it is the circle that is red, not the square.

213 The ability to also code conjunctive features, such as 'blue+square', is a strategy designed to  
214 address this issue (see e.g., Wagner & Brandon, 2000, and references therein). Like the FEAT model  
215 described above, this conjunctive feature model (**C-FEAT**) can form multiple matches, such that  
216 basic features and conjunctions in one stimulus can match with multiple features/conjunctions in  
217 the other. While such an approach is sufficient for distinguishing comparisons where features have  
218 swapped across objects (i.e., comparison *C* vs. comparison *B* in Table 1), such models cannot  
219 distinguish between identical items (comparison *A*) and a 'swap' (comparison *B*), unless they also  
220 code for relative spatial location of features. Clearly the relative spatial location of features is  
221 relevant – at least for humans (Hodgetts et al., 2009). To capture this, we can specify a new  
222 conjunctive *spatial* feature model (**CS-FEAT**), in which relative spatial features are also added, both  
223 as single features but also as components of more extended *conjunctive* features (e.g., 'left',  
224 'blue+left', and 'blue+square+left' get added to our feature set or vector) (e.g., see George et al.,  
225 2001; George & Pearce, 2003). This new model (CS-FEAT) now distinguishes comparisons *A* and *B*  
226 (i.e., identity and swap) at the expense of over-matching 'down the road' (see Table 1). Namely, the  
227 predicted similarities of comparisons *D* and *E* are now greater than *B*. Limiting feature matches to  
228 certain spatial locations (i.e., 'blue' matches 'blue' *if and only if* blue is in a corresponding spatial  
229 position) via a spatial-feature model (**S-FEAT**) seems plausible to address this, but renders equally  
230 *dissimilar* comparisons *B*, *F*, and even *G*! These stimulus materials thus draw out the fundamental  
231 limitation of features in capturing relational information. Specifically, tweaking the feature set to  
232 include conjunctive and/or relational features may seem *locally plausible* when considering just one  
233 or two comparisons, but can be shown to be *globally implausible* over an entire set of items.

234 The final two columns in Table 1 are structural models of similarity that have been applied  
235 extensively in past research (Hodgetts et al., 2009; Larkey & Markman, 2005). The first of these  
236 models, MIP (for 'matches-in-place'), draws upon existing models of structural alignment, which  
237 have been applied to capture human similarity judgements across range of contexts, including  
238 perceptual similarity (Goldstone, 1994a; Larkey & Markman, 2005), as well as metaphor and  
239 analogy (Gentner & Markman, 1997). Critically, such models assume more structured, hierarchical

240 representations, whereby local properties or attributes (e.g., 'red') form parts of whole objects (the  
241 hat), which in turn play a specific role within the broader relational structure. The alignment process  
242 itself (see e.g., Falkenhainer et al., 1989; Gentner, 1983; Gentner & Markman, 1997) seeks to form  
243 matches that are structural consistent across the two representations, which requires that a) an  
244 element in one representation must match with at most one element in the other representation  
245 (*one-to-one mapping*), and b) that wherever relations are placed in correspondence, their respective  
246 arguments are also placed in correspondence (*parallel connectivity*) (Falkenhainer et al., 1989).

247 Underpinning these structural alignment models, as demonstrated in the 'hat' and 'coat'  
248 example above, is the classic distinction between matches for elements that have been placed in  
249 correspondence (matches-in-place, or MIPs) and matches for elements that do not correspond  
250 (matches-out-of-place, or MOPs). Given the established impact of MIPs on similarity ratings in  
251 humans (Markman & Gentner, 1996), including within the stimulus domain used here (Hodgetts et  
252 al., 2009), our alignment model was based on the number of MIPs. Consistent with the *one-to-one*  
253 *mapping* constraint, a feature match on our MIP model only "counts" if the objects themselves have  
254 been placed in correspondence<sup>1</sup>. For comparison C, for instance, the lower-level features 'blue' and  
255 'square' (which make up the left-hand *object* in the target stimulus) map on to two separate *objects*  
256 in the right stimulus. Given the *one-to-one mapping* constraint, such many-to-one mappings, where  
257 a single feature/object in one stimulus is matched with two or more features/objects in a second  
258 stimulus, are not permitted under this MIP model. As a result, only a single match is counted – e.g.,  
259 the blue square is matched with the blue circle on the basis of colour, and likewise the red circle is  
260 matched with only one object in the right stimulus: the 'red square' based on the shared property  
261 'red'. This adherence to one-to-one mappings at the level of objects allows the MIP model to  
262 distinguish comparisons B and C, and also does not lead to the profound proliferation of features for  
263 other comparisons in the set.

264 The transformational model of similarity, or 'Representational Distortion' (RD) (Hahn et al.,  
265 2003), proposes that perceived similarity emerges from the complexity, or 'effort', to transform the  
266 mental representation of one object or event into another. Measures of transformational  
267 complexity, and thus (dis)similarity, may range from continuous spatial transformations (e.g.,  
268 translation, rotation, etc), as seen in models of visual object recognition (Graf, 2006; Hahn et al.,  
269 2009; Lawson, 1999; Lawson & Jolicoeur, 2003), to sets of simple operations (insert, delete, swap,

---

<sup>1</sup> Note, that different models of structural alignment do differ in their adherence to the *one-to-one mapping* constraint. For instance, while this is a strict model constraint within the Structure Mapping Engine (Gentner, 1983; Larkey & Markman, 2005), the 'Similarity, Interactive Activation and Mapping' (SIAM) model (Goldstone, 1994a) allows for differing degrees of correspondence. This means that both MIPs and MOPs influence similarity in SIAM, though MIPs still receive greater weight overall. As our prior work showed a limited influence of MOPs in this stimulus domain, and very little difference between a MIP model and SIAM (Hodgetts et al., 2009), we likewise adopted a MIP-based implementation of structural alignment in this paper.



270 etc), which can then be combined into longer codes to capture more complex transformational  
271 relationships (Hahn et al., 2003; Hahn & Bailey, 2005; Hodgetts et al., 2009). Conceptually, a single  
272 transformation may act upon individual features (or indeed continuous feature dimensions), whole  
273 ensembles of features, or manipulate the interrelationship between features or objects (i.e.,  
274 structure) (Hahn et al., 2003). In past empirical work, it has been shown that transformational model  
275 predictions can capture accurately human similarity ratings (Hahn et al., 2003; Hodgetts et al.,  
276 2009), speeded same-different judgements (Hodgetts & Hahn, 2012), and even analogical  
277 reasoning (Leech et al., 2007). Critically, it has also been shown that transformations provide  
278 superior fits of human similarity data when compared to basic feature models (Hahn et al., 2003;  
279 Toussaint et al., 2012) and models of structural alignment (Hodgetts et al., 2009; see also Larkey &  
280 Markman, 2005).

281 The RD predictions in this study (Table 1) are derived from a simple coding scheme used  
282 previously (Hodgetts et al., 2009; Hodgetts & Hahn, 2012), which been shown to capture accurately  
283 *human* perceived similarity within this stimulus domain. This coding scheme specifies three simple  
284 operations – create, apply and swap – which can be combined to characterise the transformational  
285 relationships between the items in our stimulus set (Table 1). Transformational ‘complexity’ is then  
286 operationalised in this model by the number of such transformations required by the shortest  
287 distance conversion of one object’s representation into that of another. This model, by assuming  
288 swap-like operations, can distinguish between identity (comparison *A*) and spatial changes that act  
289 upon the same set of features (i.e., comparisons *B* and *C*; Table 1). An in-detail specification of how  
290 the predictions of Table 1 are derived not just for RD, but for all of the models, can be found in the  
291 Supplementary Methods.

292 As noted in Hodgetts et al. (2009), structural alignment models and RD are not necessarily  
293 in conflict, and in many cases the preferred alignment between two object representations will be  
294 that which affords the simplest transformation between those representations (Graf, 2006).  
295 Likewise, transformational and featural, or indeed spatial models (e.g., Shepard, 1957), are not  
296 necessarily in conflict in the sense that the former can be seen as generalizations of at least some  
297 featural or spatial models – generalizations that allow a broader range of ‘transformations’  
298 including, crucially, ones that are sensitive to structure in ways that featural or spatial models are  
299 not.

300 For the purposes of the present investigation, what matters is that both structural  
301 alignment models and RD have been used successfully to provide experimental evidence for the  
302 importance of structure in human similarity judgments (e.g., Markman & Gentner, 1996; Hahn et al.,  
303 2003; Hodgetts et al., 2009; Toussaint et al., 2012). In other words, our interest is not on which of  
304 these models might be ‘best’. Rather, we will use these models as a collective set of tools for

305 probing the role of structure in perceived similarity for nonhuman primates, and thus for the role of  
306 structure in nonhuman object representation. In short, this paper seeks to probe whether there is  
307 evidence for representational schemes that go beyond mere features, while remaining agnostic to  
308 the specific ways in which such structural information might be encoded in the brain (for a selection  
309 of accounts, see e.g., Falkenhainer et al., 1989; Goldstone, 1994a; Taylor & Hummel, 2009).

310

### 311 **1.3 The current investigation**

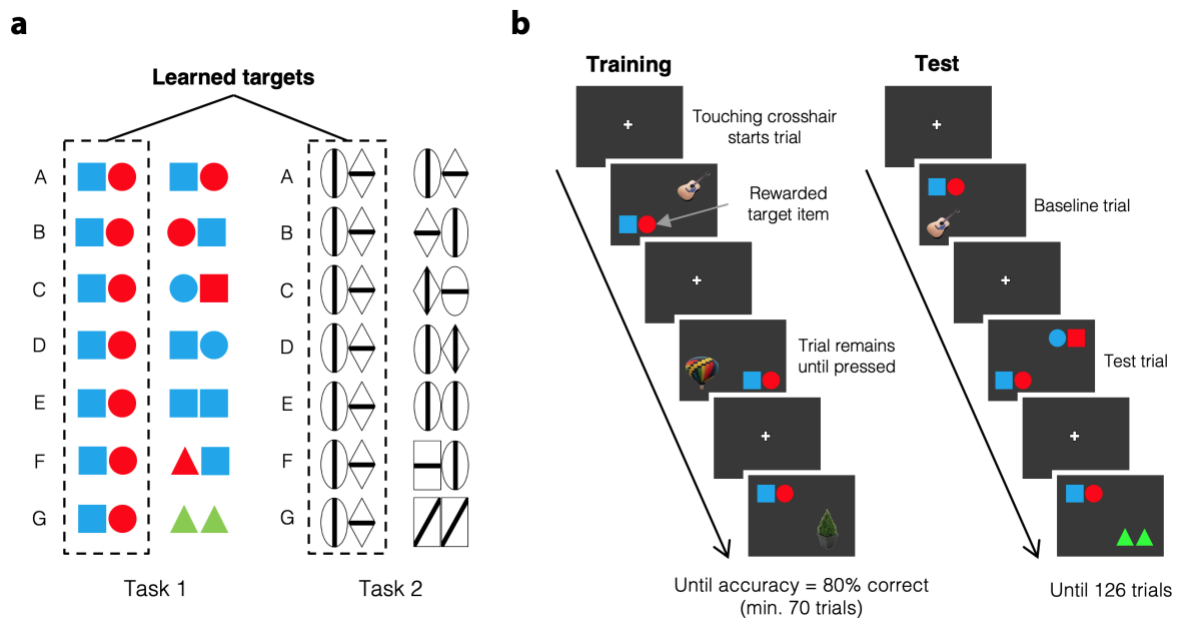
312 As highlighted in the previous section, one of the key challenges when contrasting featural and  
313 structural models of similarity is that in many contexts it is possible for featural models to 'mimic'  
314 the predictions of structural models, particularly by *ad hoc* turning structural information (e.g.,  
315 information about bound objects and spatial position) into features through the use of increasingly  
316 complex conjunctions. This has implications not only for evaluating different models of human  
317 similarity, but also, via the intimate connection between similarity and mental representation, for  
318 understanding the nature and complexity of the underlying object representations themselves. The  
319 way to avoid this mimicry is to have a carefully designed set of stimuli that allows us to demonstrate  
320 that particular *ad hoc* features, which may be effective at the level of individual comparisons, lead to  
321 counterintuitive distortions across the whole set of comparisons. In this study, we have designed  
322 such a stimulus domain, which will allow us to disentangle featural and structural models of  
323 similarity. By addressing the *ad hoc* mimicry of feature-based coding, we can provide much stronger  
324 evidence for structure *sensitivity* in both human and nonhuman species – drawing out general  
325 dichotomies between featural and structural models of similarity and their implications for  
326 understanding human and nonhuman cognition.

327 To allow direct comparisons between ape species, each species group (human, chimpanzee,  
328 and gorilla) completed the same basic tasks, where subjects had to press a specific target stimulus  
329 from two possible items on each trial (Figure 2). The difficulty of doing this (as indicated by higher  
330 error rates in nonhuman animals and slower response times in humans) was assumed to be related  
331 to higher similarity between the target and the seven test items. We constructed two versions of  
332 the stimuli (labelled Task 1 and Task 2), each with same underlying logical structure, but replacing  
333 the surface features of shape and colour with shape and the orientation of an inner line (see Figure  
334 2A). Colour, in particular, may be a core property underlying visual object discrimination and  
335 individuation in both human and nonhuman primate species (Gershkoff-Stowe & Smith, 2004;  
336 Mendes et al., 2011). Thus, we sought to compare similarities for coloured objects (Task 1) with  
337 achromatic, single-colour stimuli that manipulated only shape-related information (outer  
338 shape/inner line). Our main question, however, was whether featural models - which seem to make  
339 counter-intuitive predictions from the perspective of human observers - do, in fact, capture

340 similarity in nonhuman hominids.

341

342



343

344 **Figure 2.** (A) The stimulus comparisons used in Task 1 (left) and Task 2 (right). Each task has a target item  
345 comprising features A and B on dimension one (Task 1 = colour [blue/red]; Task 2 = line orientation  
346 [vertical/horizontal]) and A and B on dimension two (Task 1 = shape [square/circle]; Task 2 = shape  
347 [oval/diamond]). Several combinations of these features comprised the comparison stimuli, as outlined in the  
348 Methods. (B) Trial schematic for the nonhuman ape training session is shown on the left. This phase  
349 established their preference for the target items shown in panel A. Subjects received a food reward (grape)  
350 for selecting the target item over the everyday object stimulus. Training terminated when subjects reached the  
351 required criterion (80% correct). In the main experiment, which was the near-identical for human and  
352 nonhuman participants (see Methods), baseline trials (target item vs. everyday object) were intermixed with  
353 'test trials'. For test trials, targets items were paired with one of seven stimuli from the set (which vary in their  
354 similarity to the target).  
355

356

357

## 357 2 Methods

358

### 358 2.1 Participants

359

359 Five chimpanzees (*Pan troglodytes*; 1 male, 4 female) and three gorillas (*Gorilla gorilla*; 3 females)  
360 took part in Task 1 (8 apes in total). One chimpanzee (Trudi) did not complete Task 2 as she did not  
361 reach criterion (80% correct) during the learning phase, resulting in N = 4 chimpanzees in Task 2 (7  
362 apes in total). All subjects were housed at the Wolfgang Köhler Primate Research Center at Zoo  
363 Leipzig (Germany). They lived in social groups with conspecifics and had access to indoor and  
364 outdoor areas designed appropriate to the species. All subjects had touchscreen experience. Apes  
365 were tested individually within a familiar, indoor room, with the exception of one gorilla (Viringika).  
366 In Viringika's case, her young daughter accompanied her in the testing area; no obvious disruption  
367 to Viringika's performance was noted from this arrangement. Ten human volunteers (3 males, 7

368 females; mean age = 31 years; SD = 6.93) were tested at Cardiff University, School of Psychology.  
369 This was undertaken with the understanding and written consent of each participant. All  
370 participants had normal or corrected to normal vision. Informed consent was obtained after the  
371 nature and possible consequences of the study were explained to participants, in accordance with  
372 the local research ethics committee at Cardiff University. Further, the animals' care was in  
373 accordance with institutional guidelines at the Max Planck Institute for Evolutionary Anthropology,  
374 and Zoo Leipzig.

375

## 376 **2.2 Stimulus design**

377 In each task, stimuli consisted of seven test stimuli, corresponding to comparisons A-G (Table 1 &  
378 Figure 2) and seven photographs of everyday objects. Each test comparison involved two pairs of  
379 geometric shapes. Each pair was 128 x 128 pixels and individual shapes were separated by a  
380 horizontal distance of two pixels – an inter-stimulus distance that has been shown to facilitate  
381 relational processing in nonhuman primates (Fagot & Parron, 2010). The everyday object  
382 photographs were scaled to approximately the same size as the test stimuli. All stimuli were  
383 presented on a 350 x 350-pixel touch-sensitive grey square. The test stimuli were defined on two  
384 feature dimensions for each task (Task 1: Dimension 1 = 'shape', Dimension 2 = 'colour'; Task 2:  
385 Dimension 1 = 'shape', Dimension 2 = 'inner line orientation'). The value of a given stimulus on each  
386 dimension can be represented abstractly using letters, where each unique letter refers to a unique  
387 feature (Task 1 shape: A = square, B = circle, C = triangle; Task 1 colour: A = blue, B = red, C = green;  
388 Task 2 shape: A = oval, B = diamond, C = rectangle; Task 2 inner line: A = vertical, B = horizontal, C =  
389 diagonal). The arrangement of features for each comparison followed the same logical structure on  
390 both tasks. The target stimulus in each task can be denoted by AB/AB, that is, 'square to the left of  
391 circle', and 'blue to the left of red'. There were seven possible test trials corresponding to the  
392 comparisons in Figure 2A: target stimulus versus stimulus AB/AB ('identity' or same trials); target  
393 stimulus versus stimulus BA/BA; target stimulus versus stimulus BA/AB; target stimulus versus  
394 stimulus AB/AA; target stimulus versus stimulus AA/AA; target stimulus versus stimulus CA/BA;  
395 target stimulus versus stimulus CC/CC (see Figure 2A).

396

## 397 **2.3 Apparatus**

398 The nonhuman tasks were presented on a 15-inch LCD monitor, mounted in a custom-built metal  
399 holder. The monitor was situated behind a 15-inch infrared touchscreen frame. The touchscreen  
400 frame replaced one of the standard safety panels that separated the nonhuman apes from the  
401 experimenter, becoming the de facto safety panel. Situated in front of the touchscreen frame was a  
402 thin Plexiglas panel with five circular finger holes cut into it. These were in the centre, top-left, top-

403 right, bottom-left, and bottom-right corners of the panel. As well as forming an additional safety  
404 panel, the Plexiglas panel also enabled safe touching of the touchscreen frame by the apes. Located  
405 on the floor to either side of the touchscreen-metal frame, and facing towards the subjects, were  
406 two small speakers used to present audio feedback. A PC, outputting a display resolution of 1280 ×  
407 1024 pixels, was connected to the LCD monitor and touchscreen frame, and E-Prime 2.0  
408 (Psychology Software Tools, Inc., Sharpsburg, PA) was used to run the experiment. Food rewards  
409 were fed by hand through a plastic tube located next to the touchscreen-metal frame when a  
410 correct response was made. The setup was identical for the chimpanzees and the gorillas. The  
411 human version of both tasks was presented on a 13-inch laptop outputting a display resolution of  
412 1280 × 1024 pixels, and responses were made using a mouse.

413

## 414 **2.4 Non-human tasks**

### 415 **2.4.1 Task 1 - training phase**

416 When an ape entered the testing area, they received two grapes and the testing session was  
417 started. Each trial began with a white fixation point (150 × 150 pixels) presented in the centre of grey  
418 background (Figure 2B). We defined a 350 × 350-pixel touch area around the fixation point; a touch  
419 within this area initiated a 150 ms inter-stimulus interval (ISI), followed by a training trial. On each  
420 training trial, the target stimulus and one of seven everyday object stimuli was presented  
421 concurrently (Figure 2). Selection of the everyday objects on each trial was determined randomly  
422 without replacement. On each trial, the positioning of the target stimulus and picture was selected  
423 randomly from four pre-defined possibilities: target stimulus (top-left) – everyday object (top-right),  
424 and vice versa; target stimulus (bottom-left) – everyday object (bottom-right), and vice versa. A  
425 touch within the 350 × 350-pixel area of the target stimulus resulted in the immediate termination  
426 of the trial, presentation of a positive sounding sound and a food reward being given. A touch within  
427 the 350 × 350-pixel area of the real-world object stimulus resulted in the immediate termination of  
428 the trial, presentation of a negative sounding sound, no food reward and a 3000 ms delay screen as  
429 'punishment'. Following this, the same trial was repeated (i.e., a 'correction' trial) until the subject  
430 selected the rewarded target stimulus. Following a 1000 ms grey screen inter-trial interval (ITI), a  
431 new trial began with presentation of the white fixation point. Any touch made outside the pre-  
432 defined stimulus regions had no consequence. The number of trials within a session varied  
433 depending on the number of correction trials required. The minimum number of trials per training  
434 session (i.e., assuming perfect performance with no correction trials), was 70 trials. Training  
435 proceeded until the subject had achieved a response accuracy of at least 80% correct over three  
436 consecutive task runs, at which point they were transferred to the test phase.

437

#### 438 **2.4.2 Task 1 - test phase**

439 As in the training phase, the nonhuman subjects received two grapes on entering the testing area  
440 and then the touchscreen session was started. The first 21 trials of a session were identical to those  
441 shown in the training phase (i.e., the target stimulus paired with a real-world object), though no  
442 correction trials were presented (henceforth, 'baseline' trials). If an NHP chose an everyday object  
443 picture over the target stimulus, this resulted in the presentation of a 'negative' sound and no food  
444 reward being given. Following a 3000 ms delay screen ISI ('punishment'), a new trial began with  
445 presentation of the white fixation point. After completing these initial 21 trials, subjects could be  
446 presented with two trial types: 1) target stimulus everyday picture (i.e., baseline trials); and 2) target  
447 stimulus versus test items A to G (henceforth, test trials). Selection of the test stimuli was  
448 determined randomly without replacement. On each trial, the positioning of the target and test  
449 stimulus was randomly selected from the four pre-defined possibilities specified previously. In  
450 contrast to baseline trials, touching either the rewarded or unrewarded stimulus on test trials  
451 resulted in no sound stimulus being played and no food reward being given. Any touch made  
452 outside stimulus touch areas has no consequence. Following a randomly determined grey screen ITI  
453 of between 1000 ms and 3000 ms, a new trial began with presentation of the white fixation point.  
454 One block consisted of 14 baseline trials and seven test trials (presented randomly), and each run of  
455 the task consisted of five blocks. Eight runs of the task were completed in total during the test  
456 phase.

457

#### 458 **2.4.3 Task 2 - training phase**

459 The same procedure as Task 1 was used with the following exceptions: 1) a trial began with a black  
460 fixation point (150 × 150 pixels; 2) the background colour of the screen was always white; and 3) the  
461 screen positioning of the target stimulus and real-world object was pseudorandom, with the  
462 constraint that the target stimulus could not be presented on the left or right of the real-world  
463 object more than two times in a row. While there was no evidence of the nonhuman apes  
464 developing a side bias in Task 1, this control was added to reduce the possibility of a side bias being  
465 developed in Task 2.

466

#### 467 **2.4.4 Task 2 - test phase**

468 The same procedure as Task 1 was used, with the exceptions noted above and one further  
469 exception: only a maximum of four (unrewarded) test trials could be presented in a row. The reason  
470 that this was implemented was because it was possible for up to 14 unrewarded test trials to appear  
471 consecutively in Task 1 (though this never occurred). Based on this change to the experimental  
472 design, one block in Task 2 consisted of seven test trials – as in Task 1 – and between nine and 12

473 non-test trials (i.e., target shape versus everyday object picture). In each session, six blocks were  
474 completed. Overall, each session consisted of 126 trials (as in Task 1): 84 baseline trials (including  
475 the initial 21 trials) and 42 test trials.

476 Two chimpanzees (Alex and Jahaga) experienced a drop in performance on the baseline  
477 trials during the test phase. When this occurred, a correction stage was introduced whereby an  
478 incorrect response resulted in the presentation of a negative sounding sound, no food reward being  
479 given and, following a 3000 ms white screen 'punishment', a repeat of the same trial. As in the  
480 training phase, such correction continued until the subject chose the target item. Both chimpanzees  
481 that experienced the aforementioned drop in performance undertook two sessions of this modified  
482 procedure as soon as the issue arose, which quickly restored good performance on baseline trials.

483

## 484 **2.5 Human tasks**

485 Within each testing session, human participants completed both Task 1 and Task 2, and the order of  
486 these was counterbalanced between participants. Unlike the nonhuman ape version of the task,  
487 there were no training phases, each task began with 7 baseline trials (versus the 21 shown in the  
488 nonhuman version) and each task comprised a single run (versus the 8 presented in the nonhuman  
489 version). Within a single run, the task structure and parameters were identical to the nonhuman  
490 primate version.

491

## 492 **3 Results**

### 493 **3.1 Structure matters for human and nonhuman species**

494 We first examined the relationship between perceived similarity in each ape species and the  
495 predictions of our featural and structural similarity models. Assuming a power law relationship  
496 between similarity and response time (Cohen & Nosofsky, 2000; Hodgetts & Hahn, 2012), the best  
497 fitting similarity model for human participants is a structural model, RD, across both tasks; this  
498 model captures 98% of the variance for Task 1, and 90% of the variance in Task 2 (Table 2). For the  
499 nonhuman apes, the best fitting model is again RD for Task 1, capturing 58% of the variance, but for  
500 Task 2 it is outperformed by the CS-FEAT model, with 98% of the variance (the fits for each tested  
501 model can be found in Table 2).

502 This suggests that chimpanzees and gorillas processed the Task 2 materials, with its more  
503 artificial, shape-related dimensions (outer shape/inner line orientation), in a fundamentally different  
504 way to the more ecologically familiar shape-colour combination found in Task 1. In other words,  
505 there is a task difference on the basis of materials for chimpanzees and gorillas. While their  
506 performance is aligned to humans on one task, showing evidence for structure, their performance

507 looks “featural” on Task 2. In the remainder we provide analyses that follows up on this point, while  
 508 further examining the putative role of structure in the perceived similarity of nonhuman apes.

509  
 510

Default model	Task 1				Task 2			
	Human		NHA		Human		NHA	
	Linear R <sup>2</sup>	Power R <sup>2</sup>	Linear R <sup>2</sup>	Power R <sup>2</sup>	Linear R <sup>2</sup>	Power R <sup>2</sup>	Linear R <sup>2</sup>	Power R <sup>2</sup>
FEAT	0.33	0.39	0.21	0.17	0.41	0.47	0.66	0.53
C-FEAT	0.33	0.38	0.16	0.15	0.32	0.42	0.67	0.55
CS-FEAT	0.43	0.43	0.25	0.22	0.29	0.41	<b>0.98*</b>	0.78
S-FEAT	0.32	0.22	0.23	0.21	0.17	0.11	0.81	<b>0.80*</b>
CMOD-SM	0.60	0.59	0.45	0.15	0.76	0.71	0.53	0.48
MIP	0.57	0.54	0.30	0.23	0.66	0.65	0.43	0.45
RD	<b>0.73*</b>	<b>0.98*</b>	<b>0.58*</b>	<b>0.80*</b>	<b>0.85*</b>	<b>0.88*</b>	0.57	0.57

Weighted model	Task 1				Task 2			
	Human		NHA		Human		NHA	
	Linear R <sup>2</sup>	Power R <sup>2</sup>	Linear R <sup>2</sup>	Power R <sup>2</sup>	Linear R <sup>2</sup>	Power R <sup>2</sup>	Linear R <sup>2</sup>	Power R <sup>2</sup>
FEAT	0.36	0.43	0.23	0.20	0.42	0.47	0.74	0.63
C-FEAT	0.36	0.43	0.23	0.20	0.42	0.47	0.74	0.63
CS-FEAT	0.45	0.46	0.44	0.32	0.45	0.49	<b>0.97*</b>	<b>0.88*</b>
S-FEAT	0.37	0.28	0.46	0.38	0.25	0.17	0.81	0.83
CMOD-SM	0.63	0.68	0.52	0.41	0.80	0.79	0.64	0.60
MIP	0.53	0.58	0.53	0.40	0.77	0.77	0.20	0.26
RD	<b>0.75*</b>	<b>0.92*</b>	<b>0.70*</b>	<b>0.87*</b>	<b>0.85*</b>	<b>0.88*</b>	0.79	0.78

511  
 512 **Table 2.** Goodness of fit (R<sup>2</sup>) for each similarity model tested. The fits for the default models (i.e., the  
 513 predictions shown in Table 1) are shown in the upper table. The lower table displays the model fits when the  
 514 relative weighting of each stimulus dimension (e.g., colour and shape) is allowed to vary parametrically (see  
 515 Section 3.4 for further detail). The fits of the Contrast Model are also reported (labelled CMOD-SM; see  
 516 Section 3.5 and Supplementary Methods). Both linear and power fits are shown, and models are fitted to  
 517 response time data for the human participants and accuracy for the nonhuman ape (NHA) species. Prior work  
 518 has shown that similarity-response time relationships are readily captured by a power law (Cohen & Nosofsky,  
 519 2000; Hodgetts & Hahn, 2012), and so power fits are used when assessing model fits of human data.

520  
 521

### 522 3.2 Task differences emerge in nonhuman apes but not humans

523 The difference between these tasks, as well as an across-species structure *sensitivity* in Task 1, is  
 524 confirmed by a global analysis that combines all the goodness of fit values (R-squared) for the  
 525 ‘featural’ models found in Table 1 into a single feature measure, and likewise all ‘structural’ models  
 526 and their goodness of fit into a structure measure. To derive a simple metric of ‘structurality’ from  
 527 this model space, we calculated a difference score between the structure and feature measures for

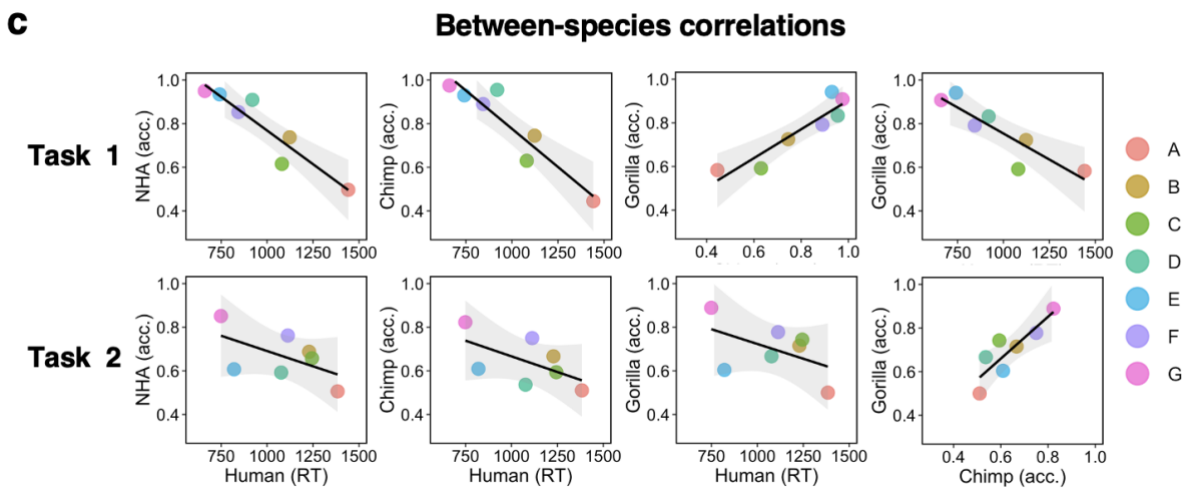
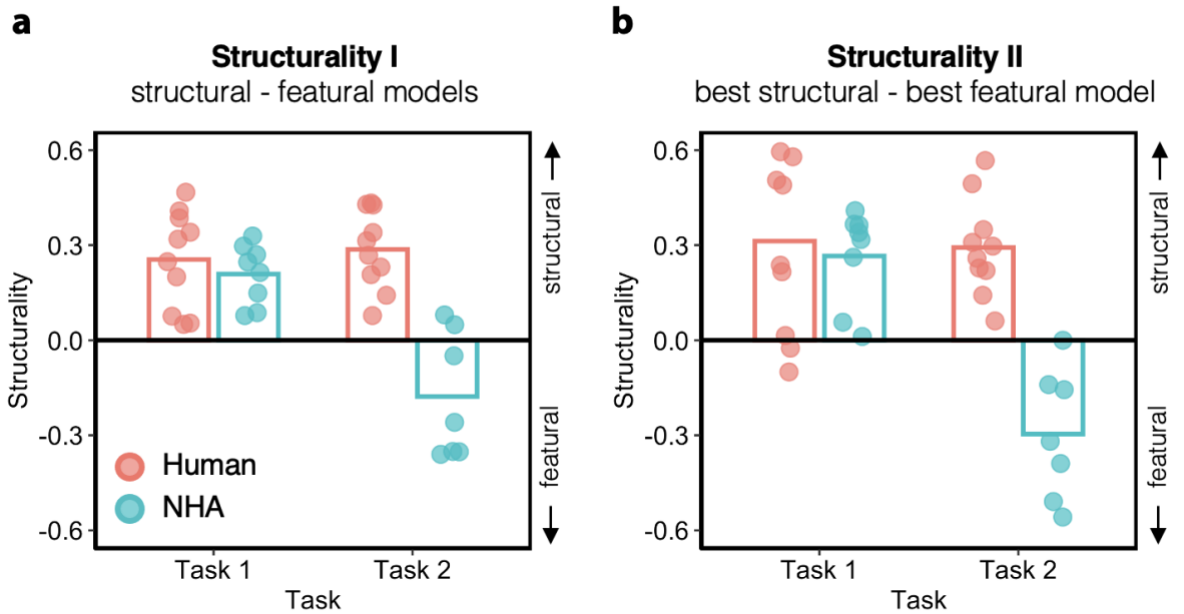


528 each individual subject. A positive score on this metric indicates that the pattern of perceived  
529 similarity across comparisons – at an *individual-level* – is better captured by structural models of  
530 similarity.

531 Consistent with previous studies of both similarity and analogy (Hahn et al., 2003; Hodgetts  
532 et al., 2009; Markman & Gentner, 1993), all human participants (10/10) were more structural (Figure  
533 3A). Strikingly, the same pattern was also observed for the nonhuman ape species, with all  
534 individual subjects scoring positively on the structurality measure and exhibiting scores that appear  
535 in the centre of human structurality distribution in Task 1 (Figure 3A). This shows, therefore, that  
536 the structural models accounted for more of the variance in individual-level similarity data than  
537 featural models across *all* human and nonhuman apes.

538 When this metric was compared for Task 2 – in which human and nonhuman apes were  
539 weakly correlated in terms of general patterns of similarity (Figure 3C; Supplementary Results) – the  
540 apes appear more featural overall (Figure 3A). The humans, however, are numerically even more  
541 structural in Task 2 (Task 1 mean = 0.26; Task 2 mean = 0.29). These observations were confirmed  
542 statistically, such that the structurality difference between human and nonhuman ape species was  
543 strongly dependent on task stimuli ( $F(1, 15) = 14.99, p = 0.002, \eta_p^2 = 0.5, BF_{10} = 2793.9$ ; Figure 3A).  
544 Follow-up Welch t-tests revealed no significant difference in the structurality measure in Task 1 ( $p =$   
545  $0.45; BF_{10} = 0.55$ ) but a significant difference in Task 2 ( $p < 0.001; BF_{10} = 594.6$ ). Furthermore, while  
546 the human structurality scores did not differ between tasks ( $p = 0.62; BF_{10} = 0.35$ ), the nonhuman  
547 ape group were significantly more 'featural' in Task 2 ( $p = 0.005; BF_{10} = 10.82$ ). The same interaction  
548 emerges when considering the single best-fitting similarity model for each subject (Table S1). Here,  
549 the best fitting model is structural for the majority of the human participants in both Task 1 (8/10)  
550 and Task 2 (10/10), and for all chimpanzees and gorillas in Task 1 (8/8). The featural models,  
551 however, provide the best fit for 5/7 nonhuman apes in Task 2 (Table S1).

552 As the structurality measure includes a larger pool of featural models, it may underestimate  
553 the role of features by also considering those which provide a poor fit of the data in a given subject.  
554 To address this possibility, we recalculated the structurality measure in each subject (across species)  
555 and subtracted the single best fitting feature model from the best fitting structural model (see  
556 Figure 3B). The same interaction was observed ( $F(1, 15) = 18.17, p = 0.001, \eta_p^2 = 0.46, BF_{10} = 1821.27$ ;  
557 Figure 3B), such that no significant difference was seen between groups in Task 1 ( $p = 0.65; BF_{10} =$   
558  $0.47$ ), but a difference was found in Task 2 ( $p < 0.001; BF_{10} = 2129.69$ ).



559

560 **Figure 3.** Sensitivity to structure in both human and nonhuman ape species (labelled 'NHA' in the figure). (A)  
 561 'Structurality' scores were calculated by fitting each model to similarity data in each individual subject  
 562 (response time (RT) for humans and accuracy (acc.) for nonhumans) and subtracting the mean fit (R-squared)  
 563 of the featural models from the mean fit of structural models (see main text). Mean structurality scores are  
 564 shown for both humans (red bars) and nonhuman apes (cyan bars) for both Task 1 (left) and Task 2 (right). A  
 565 positive score indicates that an individual's data is fit better by structural (MIP, RD) versus featural models  
 566 (FEAT, C-FEAT, CS-FEAT, S-FEAT). Individual data points are shown for each subject in Task 1 (Human: n =  
 567 10; Non-human: n = 8) and Task 2 (Human: n = 10; Non-human: n = 7). (C) Correlations between human and  
 568 nonhuman ape similarity judgements for Task 1 and Task 2. From left to right, the graphs depict the  
 569 correlation between human and nonhuman apes (left), humans and chimpanzees (left-middle), humans and  
 570 gorillas (right-middle) and chimpanzees and gorillas (right). Each scatter plot contains seven data points, each  
 571 reflecting mean RT (humans) and accuracy (nonhumans) for each stimulus comparison (A to G).

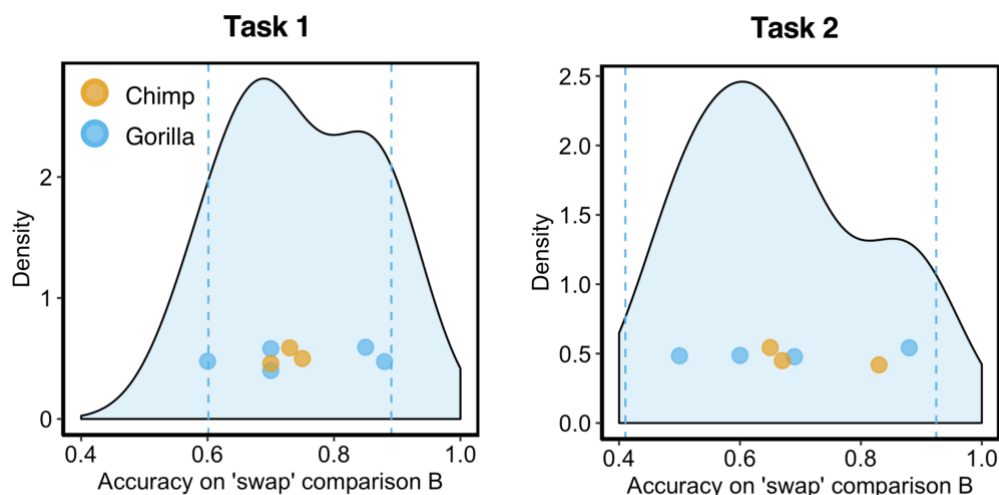
572

573

### 574 3.3 Apes are sensitive to spatial location

575 The same picture also emerges when we consider individual item comparisons. As noted in the  
 576 Introduction, the key challenge for featural models is distinguishing between identity and swap,

577 that is, stimulus comparisons *A* and *B* (Table 1) - a distinction that can readily be captured by  
 578 structural models. In a featural model, however, this can only be achieved by adding some  
 579 sensitivity to spatial location, such as the relative position of objects within the stimulus. But this  
 580 then invariably creates problems with comparisons *E*, *F* and *G* in Table 1, where the specificity  
 581 introduced by spatial features (to distinguish identity and swap) distorts the similarity across the set  
 582 as a whole. Notably, all the species groups can distinguish between identity and swap (i.e.,  
 583 comparisons *A* and *B*; Figure S3). The nonhuman apes, for instance, are exactly at chance for  
 584 comparison *A* but are significantly more accurate at identifying the target item for the 'swap'  
 585 comparison in both Task 1 ( $p = 0.002$ ;  $BF_{10} = 29.6$ ) and Task 2 ( $p = 0.007$ ,  $BF_{10} = 9.3$ ). Incidentally,  
 586 though it has been claimed that gorillas are less sensitive to structure than chimpanzees (Haun &  
 587 Call, 2009), on our items all three gorillas were able to distinguish identity and swap above chance  
 588 and indeed fall within the centre of the chimpanzee distribution on both tasks (Figure 4).  
 589  
 590



591  
 592 **Figure 4.** Gorillas and chimpanzees are similarly sensitive to structure. The accuracy for each nonhuman  
 593 subject on the 'swap' comparison *B* is shown for (A) Task 1 and (B) Task 2. The gorillas are shown to be within  
 594 the chimpanzee distribution on both tasks. The density distribution and 95% confidence interval for the  
 595 chimpanzees is shown in blue. The individual data points for the chimpanzees and gorillas are depicted by  
 596 blue and orange markers, respectively.  
 597

598  
 599 For humans, the pair of stimuli in comparison *E* are highly dissimilar in both tasks (Figure S3), but  
 600 the feature models with spatial sensitivity (see 'S-FEAT' and 'CS-FEAT' in Table 1) fail to capture this  
 601 by rendering *E* more similar than *B* (e.g., the swap). This limitation of the feature model, however, is  
 602 relevant to performance in the nonhuman group; while the stimuli in comparison *E* are perceived to  
 603 be dissimilar in Task 1, 3/7 nonhuman apes found the items in *E* to be as highly similar as in identical

604 comparison *A*! This indicates difficulty in accurately encoding relational information for stimuli  
605 based purely on line information, giving rise to the interactions observed in Figure 3. Note, however,  
606 that these subjects' behavioural responses to *E* do not reflect a simple 'blindness' to one of the two  
607 feature dimensions in task 2 (i.e., shape or line orientation), as we detail next.

608

### 609 **3.4 Species-specific differences in dimensional salience**

610 The relevance of specific feature dimensions to perceived similarity can be assessed by deriving  
611 model predictions for each dimension separately and then giving them different weight in  
612 determining similarity (100% colour, 90% colour/10% shape, and so on; see (Hodgetts et al., 2009).  
613 This allows us to examine the performance of models for dimension 1 only, dimension 2 only, and all  
614 relative weightings in between (see Table 2 and Figure S1A). Exploration of the best-fitting  
615 dimension weights for the well-performing models suggests that humans are sensitive to both  
616 colour and shape in Task 1, whereas chimpanzees and gorillas are, if anything, slightly more  
617 sensitive to colour — an observation that resonates with previous findings (Mendes et al., 2011).  
618 Importantly, the same analyses for Task 2 suggest that, in contrast to human subjects, the  
619 nonhuman ape subjects struggled to incorporate the orientation of the inner line of the Task 2  
620 stimulus (Figure S1A). This finding is also confirmed in a model-independent analysis that examines  
621 patterns of equivalence between comparisons that would be obtained if participants were blind to  
622 one or both of the dimensions (see Supplementary Results and Figure S1B).

623 Note, however, that a participant's focus on a single dimension only could still provide  
624 evidence for structure-sensitivity. For instance, human experimental studies that applied similar  
625 stimuli to evaluate a range of structural models considered variation in only a single dimension  
626 (Hodgetts et al., 2009). Likewise in the present materials, the qualitative difference in respect to  
627 comparison *E* across Task 1 and Task 2 for nonhuman apes cannot be explained in reference to  
628 selective dimension blindness, as comparison *E* is distinct from *A* regardless of whether a  
629 participant is paying attention to both dimensions, or a single one (Figure S2). Finally, it is worth  
630 noting that even in Task 2 both dimensions (inner line/shape) mattered to chimpanzees and gorillas,  
631 as seen from the fact that there is anecdotal evidence for a difference between the judged similarity  
632 of comparisons *A* and *D*, which are identical in respect to the more salient outer shape dimension ( $t$   
633 = 2.12,  $p = 0.08$ ;  $BF_{10} = 1.4$ ); this suggests some sensitivity to the inner line.

634

### 635 **3.5 Common and distinctive features**

636 Finally, it is worth commenting on the performance of Tversky's Contrast Model (Tversky, 1977).  
637 While virtually unknown in animal literature, it is the most well-known featural model in the human  
638 similarity literature. The crucial feature of the Contrast Model is that similarity is a function of both

639 the common and the unique features of each object in a comparison. Unlike the featural models  
640 outlined in Table 1, therefore, it does not just look for commonalities but also factors in those  
641 feature or attributes that distinguish objects (see Supplementary Methods for a full description of  
642 the model, and Table 1 and Table S2 for overall and individual-level fits, respectively). Given the  
643 nature of our stimuli, the difference between counting just matches, and counting matches and  
644 mismatches is more apparent than real: the key value of Tversky's Contrast Model is realised in  
645 situations where object descriptions differ in complexity such that the size (and weight) of the sets  
646 of distinctive features of each object vary. This, for example, is what allows the model to capture  
647 asymmetries in the context of directional similarity judgments. But these conditions are not met in  
648 our stimulus materials: every stimulus item consists of exactly two components, each of which has a  
649 colour and a shape. The number of matches is thus taken from the same wider set of possibilities in  
650 each case, so that enumerating the matching features effectively identifies the complement set of  
651 the non-matching features as well. As a result, we would not necessarily expect this model to  
652 perform vastly better than other featural models for our stimulus domain, even if contrasting  
653 features were relevant to all our participants. Indeed, while the Contrast Model is the best  
654 performing *featural* model for the humans on both Task 1 and Task 2, it is not the best featural  
655 model for nonhuman subjects (see CMOD-SM in Table 1). In fact, no single featural model performs  
656 best across Task 1 and Task 2 for the nonhuman apes; a finding that obtains both overall *and* at the  
657 individual level (see Tables S1-2). Similar to the individual-level fits reported in Section 3.2, the best  
658 fitting model is structural for the majority of human participants in Task 1 (7/10) and Task 2 (10/10)  
659 when the Contrast Model is also considered (Table S2). For nonhuman apes, the Contrast Model  
660 does not provide the best fit in any single subject for Task 1 or Task2 (Table S2). Finally, the key  
661 interaction between group and task in Figure 3 is retained when the Contrast Model is included in  
662 the structurality scores ( $F(1, 15) = 12.8, p = 0.003, \eta_p^2 = 0.46; BF = 1164.935$ ).

663

#### 664 **4 Discussion**

665 Contemporary models of human similarity place critical emphasis on the role of 'structure' in the  
666 representation of real-world objects and events (Gentner & Markman, 1997; Hahn et al., 2003),  
667 meaning that both individual elements and, critically, how they are interrelated can influence  
668 perceived similarity. Despite the relevance of structural information to human perceptual similarity,  
669 object recognition, analogical reasoning, and so on, dominant models of animal learning and  
670 cognition (e.g., (Pearce, 1994; Rescorla & Wagner, 1972) still assume that nonhuman similarity and  
671 generalisation can be readily captured by simple, feature-based representation schemes.  
672 Challenging this view, we have systematically demonstrated in this study that structural models of  
673 similarity can precisely capture, like in adult humans, patterns of similarity judgements in

674 nonhuman species, specifically great apes (chimpanzees and gorillas). By applying a novel measure  
675 of 'structurality' – derived by contrasting a set of featural models with a set of established structural  
676 models of similarity – we found that all nonhuman apes demonstrated sensitivity to structure when  
677 discriminating basic object stimuli (coloured shapes). This provides strong evidence that nonhuman  
678 ape species utilise complex object representations that go beyond basic feature sets.

679       Importantly, the claim that similarity comparisons in nonhuman apes seem to involve  
680 structured representations is distinct from the question of whether they can or cannot recognise  
681 *purely relational* similarities. Prior work has shown that many nonhuman species have difficulty  
682 recognising similarities that are based entirely on abstract relational properties, such as identifying  
683 that two squares and two circles share the relational property of "sameness" (Blough, 2001; George  
684 et al., 2001; Haun & Call, 2009); see also (Penn et al., 2008; Tomasello & Call, 2007). Obviously, if  
685 relational information is not even encoded, purely relational matches are impossible; even where  
686 relational information is encoded, however, it may not support arbitrary comparisons. Purely  
687 relational properties such as "sameness" are highly abstract, higher-level aspects of comparisons;  
688 that these may be challenging is entirely in keeping with our own results indicating that building  
689 structured representations out of stimuli composed of arbitrary feature combinations is already  
690 difficult for nonhuman primates.

691       While all nonhuman apes demonstrated a striking sensitivity to structure for coloured  
692 shapes, they were notably more *featural* in Task 2, which involved line information only (inner and  
693 outer contours). Additional model-based and model-free analyses suggested that nonhuman apes  
694 had particular difficulty integrating the outer shape with the inner line orientation. This interesting  
695 result aligns with prior work that highlights a central role for colour (versus shape) information in  
696 nonhuman primate object individuation and discrimination (Mendes et al., 2011; Santos et al.,  
697 2002). While past cross-species work has found some congruence with human results with respect  
698 to the feature binding of colour and shape (e.g., Buračas & Albright, 1999; Cook, 1992)<sup>2</sup>, research  
699 with other types of 'features' (such as the distinction between local and global features) has  
700 suggested the possibility of significant cross-species differences (e.g., (Fagot & Tomonaga, 1999;  
701 Hopkins & Washburn, 2002) in how attributes of compound stimuli are perceived. In the context of  
702 our central question about the representation of relational structure, the contrast between  
703 performance in Task 1 and Task 2 highlights how the functional relevance of different forms of  
704 lower-level stimulus information may impact on the expression of higher-order cognitive processes.  
705 For example, while colour was highly relevant in this visual discrimination paradigm, it may be that  
706 shape information would be weighted more in contexts where it is ecologically/functionally

---

<sup>2</sup> Note, this does not mean that nonhuman species can necessarily use these object representations in the same way that humans can (e.g., see Smith et al., 2004)

707 relevant, such as when selecting stimuli for action, as seen in tool selection (Santos et al., 2003).  
708 Importantly, while the nonhuman apes failed to show structural sensitivity to the line-only stimuli of  
709 Task 2, they did show evidence of attending to both 'dimensions' (outline shape/inner line) in their  
710 discrimination. This suggests that the difference between the two tasks may reflect a performance  
711 difficulty rather than an in-principle limitation. Specifically, the highly salient, accessible, and  
712 familiar dimension of colour may be easier to process, leaving greater resources for the integration  
713 of the second dimension (even human participants were somewhat slower in Task 2, hinting at a  
714 visual processing advantage for the coloured objects in Task 1). In other words, the task differences  
715 for the nonhuman apes may reflect the kinds of performance differences seen widely for humans in  
716 cognitive and developmental psychology across familiar and less familiar materials. It thus seems  
717 possible that greater sensitivity to structure in Task 2 would emerge after more extensive training  
718 with the basic discrimination. This dissociation suggests also that requiring the binding mechanism  
719 to be wholly independent of the elements to-be-bound (as assumed to be definitional of structured  
720 representation in (Doumas & Martin, 2018, p. 169) may be too strong a constraint for the meaning  
721 of structure in nonhuman species. And indeed our results might be taken to suggest that it would be  
722 empirically and conceptually fruitful to not take the issue of structure in nonhuman primates as all  
723 or none (Doumas & Martin, 2018).

724 That said, the results from Task 1 do mean, however, that models that do not include  
725 relational information are too restrictive for nonhuman apes, at least some of the time. Featural or  
726 vector-based models of stimulus representation, and as a consequence similarity, continue to  
727 dominate cognitive psychology (Ashby, 1992; Galesic et al., 2018; Hout et al., 2014), animal learning  
728 (Hall & Rodríguez, 2017; Holmes et al., 2019; Luzzardo et al., 2017; Pearce, 2008; Rasmussen et al.,  
729 2015), cognitive neuroscience (King et al., 2019; Mur et al., 2013; Theves et al., 2019) and machine  
730 learning (Hamel, 2009). Though arguments about the empirical adequacy of such representation  
731 schemes have been repeatedly made, they continue to be popular, arguably, because it is  
732 empirically difficult to provide compelling evidence for the role of structure in determining visual  
733 similarity. For one, this stems from the difficulty of excluding 'mimicry' of seeming sensitivity to  
734 structure through featural models. However, it stems also from the idea that structural similarities  
735 may be restricted in their role to 'higher-level' contexts, such as those involving some form of  
736 analogical reasoning (Hahn, 2014). This has made it seem plausible that structure does not matter  
737 to visual similarity in most, or all, nonhuman species. The present results refute this. Relational  
738 information has often figured as a fault line between researchers who view nonhuman species as  
739 capable of high-level abilities, such as causal reasoning (Beckers et al., 2006; Blaisdell et al., 2006;  
740 Call, 2004), theory of mind (Tomasello et al., 2003), imitation (Tomasello et al., 2005) or mental  
741 time travel (Clayton & Dickinson, 2009; Martin-Ordas et al., 2013), on the one hand, and those who

742 view seeming demonstrations of such abilities as the product of simpler associative processes  
743 (Dwyer et al., 2009; Heyes, 1998, 2001; Penn & Povinelli, 2007). In this context, the present results  
744 emphasise the need to distinguish carefully between relations embedded within object and/or  
745 event representations and true higher-order relational reasoning (Penn et al., 2008). Indeed, these  
746 results emphasise the need to develop more realistic models of the former, which include relational  
747 structure, even in species that show no evidence of the latter.

748 On a methodological level, the present paper shows that effective tests for the role of  
749 structure in perceptual similarity can be derived by thwarting attempts at (plausible) mimicry  
750 through systematic expansion of the set of comparisons. Given such a test, the limitations of  
751 featural models for human perceptions of similarity emerge very clearly, but they also emerge not  
752 just for chimpanzees but also the more distantly related gorillas – at least for those types of stimuli  
753 that combine familiar object dimensions.

754

755

---

## 756 Additional information

757

### 758 Supplementary material

759 Data, task materials/stimuli, as well as R code to reproduce key analyses and figures, are freely  
760 available on the Open Science Framework: <https://osf.io/jkfdz/>

761

### 762 Acknowledgements

763 We would like to thank John Pearce, John Aggleton, Craig Hedge, Petroc Sumner and the Cardiff  
764 University School of Psychology Cognitive Group for their helpful discussion and valuable  
765 feedback. Further we would like to thank the keepers of Zoo Leipzig for their assistance in this  
766 research.

767

### 768 Funding sources

769 This work was supported by the Biotechnology and Biological Sciences Research Council  
770 [BB/V010549/1] and the European Commission New and Emerging Science and Technology (NEST)  
771 Scheme [516542].

772

### 773 CRediT author statement

774 Carl Hodgetts: Conceptualization, Investigation (human), Formal Analysis, Data Curation,  
775 Visualization, Writing – Original Draft, Writing - Reviewing and Editing; James Close:



776 Conceptualization, Investigation (nonhuman), Formal Analysis, Writing - Reviewing and Editing;  
777 Ulrike Hahn: Conceptualization, Formal Analysis, Writing – Original Draft, Writing - Reviewing and  
778 Editing.

779

780

---

## 781 References

782 Ashby, F. G. (1992). *Multidimensional models of perception and cognition* (F. G. Ashby, Ed.). Psychology Press.

783 Beckers, T., Miller, R. R., De Houwer, J., & Urushihara, K. (2006). Reasoning rats: Forward blocking in  
784 pavlovian animal conditioning is sensitive to constraints of causal inference. *Journal of Experimental*  
785 *Psychology: General*, *135*(1), 92–102. <https://doi.org/10.1037/0096-3445.135.1.92>

786 Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological*  
787 *Review*, *94*(2), 115–117. <https://doi.org/10.1037/0033-295X.94.2.115>

788 Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, *311*(5763),  
789 1020–1022. <https://doi.org/10.1126/science.1121872>

790 Blough, D. (2001). The perception of similarity. In R. G. Cook (Ed.), *Avian Visual Cognition*. Comparative  
791 Cognition Press.

792 Blumberg, M. S., & Sokoloff, G. (2001). Do infant rats cry? *Psychological Review*, *108*(1), 83–95.  
793 <https://doi.org/10.1037/0033-295X>.

794 Buračas, G. T., & Albright, T. D. (1999). Covert visual search: A comparison of performance by humans and  
795 macaques (*Macaca mulatta*). *Behavioral Neuroscience*, *113*(3), 451–464. [https://doi.org/10.1037/0735-](https://doi.org/10.1037/0735-7044.113.3.451)  
796 [7044.113.3.451](https://doi.org/10.1037/0735-7044.113.3.451)

797 Call, J. (2004). Inferences about the location of food in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla*  
798 *gorilla*, and *Pongo pygmaeus*). *Journal of Comparative Psychology*, *118*(2), 232–241.  
799 <https://doi.org/10.1037/0735-7036.118.2.232>

800 Cheries, E. W., Newman, G. E., Santos, L. R., & Scholl, B. J. (2006). Units of visual individuation in rhesus  
801 macaques: Objects or unbound features? *Perception*, *35*(8), 1057–1071. <https://doi.org/10.1068/p5551>

802 Clayton, N. S., & Dickinson, A. (2009). Mental time travel: Can animals recall the past and plan for the future?  
803 *Encyclopedia of Animal Behavior*, *4*(August), 438–442. [https://doi.org/10.1016/B978-0-08-045337-](https://doi.org/10.1016/B978-0-08-045337-8.00094-2)  
804 [8.00094-2](https://doi.org/10.1016/B978-0-08-045337-8.00094-2)

805 Cohen, A. L., & Nosofsky, R. M. (2000). An exemplar-retrieval model of speeded same–different judgments.  
806 *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1549–1569.  
807 <https://doi.org/10.1037/0096-1523.26.5.1549>

808 Cook, R. G. (1992). Dimensional organization and texture discrimination in pigeons. *Journal of Experimental*  
809 *Psychology: Animal Behavior Processes*, *18*(4), 354–363. <https://doi.org/10.1037/0097-7403.18.4.354>

810 Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society*  
811 *B: Biological Sciences*, *367*(1603), 2733–2742. <https://doi.org/10.1098/rstb.2012.0220>

812 Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of  
813 relational concepts. *Psychological Review*, *115*(1), 1–43. <https://doi.org/10.1037/0033-295X.115.1.1>

814 Dumas, L. A. A., & Martin, A. E. (2018). Learning structured representations from experience. In *Psychology*  
815 *of Learning and Motivation—Advances in Research and Theory* (1st ed., Vol. 69). Elsevier Inc.  
816 <https://doi.org/10.1016/bs.plm.2018.10.002>

817 Dwyer, D. M., Starns, J., & Honey, R. C. (2009). "Causal reasoning" in rats: A reappraisal. *Journal of*  
818 *Experimental Psychology: Animal Behavior Processes*, 35(4), 578–586.  
819 <https://doi.org/10.1037/a0015007>

820 Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, 21(4), 449–  
821 498. <https://doi.org/10.1017/S0140525X98001253>

822 Fagot, J., & Parron, C. (2010). Relational matching in baboons (*Papio papio*) with reduced grouping  
823 requirements. *Journal of Experimental Psychology: Animal Behavior Processes*, 36(2), 184–193.  
824 <https://doi.org/10.1037/a0017169>

825 Fagot, J., & Thompson, R. K. R. (2011). Generalized relational matching by guinea baboons (*Papio papio*) in  
826 two-by-two-item analogy problems. *Psychological Science*, 22(10), 1304–1309.  
827 <https://doi.org/10.1177/0956797611422916>

828 Fagot, J., & Tomonaga, M. (1999). Global and local processing in humans (*Homo sapiens*) and chimpanzees  
829 (*Pan troglodytes*): Use of a visual search task with compound stimuli. *Journal of Comparative*  
830 *Psychology*, 113(1), 3–12.

831 Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and  
832 examples. *Artificial Intelligence*, 41, 1–63. [https://doi.org/10.1207/s15327868ms0702\\_6](https://doi.org/10.1207/s15327868ms0702_6)

833 Galesic, M., Walkyria Goode, A., Wallsten, T. S., & Norman, K. L. (2018). Using Tversky's contrast model to  
834 investigate how features of similarity affect judgments of likelihood. *Judgment and Decision Making*,  
835 13(2), 163–169.

836 Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170.  
837 <https://doi.org/10.1054/arth.2001.25504>

838 Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*,  
839 52(1), 45–56. <https://doi.org/10.1037//0003-066X.52.1.45>

840 George, D. N., & Pearce, J. M. (2003). Discrimination of structure: II. Feature binding. *Journal of Experimental*  
841 *Psychology: Animal Behavior Processes*, 29(2), 107–117. <https://doi.org/10.1037/0097-7403.29.2.107>

842 George, D. N., Ward-Robinson, J., & Pearce, J. M. (2001). Discrimination of structure: I. Implications for  
843 connectionist theories of discrimination learning. *Journal of Experimental Psychology: Animal*  
844 *Behavior Processes*, 27(3), 206–218. <https://doi.org/10.1037/0097-7403.27.3.206>

845 Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75(4),  
846 1098–1114. <https://doi.org/10.1111/j.1467-8624.2004.00728.x>

847 Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model.  
848 *Journal of Experimental Psychology. General*, 117(3), 227–247. [https://doi.org/10.1037/0096-](https://doi.org/10.1037/0096-3445.117.3.227)  
849 [3445.117.3.227](https://doi.org/10.1037/0096-3445.117.3.227)

850 Goldstone, R. L. (1994a). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology:*  
851 *Learning, Memory, and Cognition*, 20(1), 3–28. <https://doi.org/10.1037/0278-7393.20.1.3>

852 Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52(2),  
853 125–157. [https://doi.org/10.1016/0010-0277\(94\)90065-5](https://doi.org/10.1016/0010-0277(94)90065-5)

854 Goldstone, R. L., & Medin, D. L. (1994). Time course of comparison. *Journal of Experimental Psychology:*  
855 *Learning, Memory, and Cognition*, 20(1), 29–50. <https://doi.org/10.1037/0278-7393.20.1.29>

856 Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin*, 132(6), 920–945.  
857 <https://doi.org/10.1037/0033-2909.132.6.920>

858 Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, 25(6), 475–492.  
859 <https://doi.org/10.1016/j.tics.2021.01.006>

860 Hahn, U. (2014). Similarity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(3), 271–280.  
861 <https://doi.org/10.1002/wcs.1282>

862 Hahn, U., & Bailey, T. M. (2005). What makes words sound similar? *Cognition*, 97(3), 227–267.  
863 <https://doi.org/10.1016/j.cognition.2004.09.006>

864 Hahn, U., & Chater, N. (1997). Concepts and similarity. *Knowledge, Concepts, and Categories (Studies in*  
865 *Cognition)*, 43–92.

866 Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, 87, 1–32.  
867 <https://doi.org/10.1016/S0>

868 Hahn, U., Close, J., & Graf, M. (2009). Transformation direction influences shape-similarity judgments.  
869 *Psychological Science*, 20(4), 447–454. <https://doi.org/10.1111/j.1467-9280.2009.02310.x>

870 Hall, G., & Rodríguez, G. (2017). Habituation and conditioning: Salience change in associative learning. *Journal*  
871 *of Experimental Psychology: Animal Learning and Cognition*, 43(1), 48–61.  
872 <https://doi.org/10.1037/xan0000129>

873 Hamel, L. (2009). Knowledge discovery with support vector machines. In *Knowledge Discovery with Support*  
874 *Vector Machines*. Wiley. <https://doi.org/10.1002/9780470503065>

875 Haun, D. B. M., & Call, J. (2009). Great apes' capacities to recognize relational similarity. *Cognition*, 110(2),  
876 147–159. <https://doi.org/10.1016/j.cognition.2008.10.012>

877 Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, 21(1), 101–148.  
878 <https://doi.org/10.1017/S0140525X98000703>

879 Heyes, C. M. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, 5(6), 253–261.

880 Hodgetts, C. J., & Hahn, U. (2012). Similarity-based asymmetries in perceptual matching. *Acta Psychologica*,  
881 139(2), 291–299. <https://doi.org/10.1016/j.actpsy.2011.12.003>

882 [dataset] Hodgetts, C. J., & Hahn, U. (2023, February 17). Similarity and structured representation in human  
883 and non-human apes. Retrieved from [osf.io/jkfdz](https://osf.io/jkfdz)

884 Hodgetts, C. J., Hahn, U., & Chater, N. (2009). Transformation and alignment in similarity. *Cognition*, 113(1),  
885 62–79. <https://doi.org/10.1016/j.cognition.2009.07.010>

886 Holmes, N. M., Chan, Y. Y., & Westbrook, R. F. (2019). A combination of common and individual error terms is  
887 not needed to explain associative changes when cues with different training histories are conditioned  
888 in compound: A review of Rescorla's compound test procedure. *Journal of Experimental Psychology.*  
889 *Animal Learning and Cognition*, 45(2), 242–256. <https://doi.org/10.1037/xan0000204>

890 Hopkins, W. D., & Washburn, D. A. (2002). Matching visual stimuli on the basis of global and local features by  
891 chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *Animal Cognition*, *5*(1), 27–31.  
892 <https://doi.org/10.1007/s10071-001-0121-8>

893 Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with  
894 similarity ratings for 240 object categories from the massive memory picture database. *PLoS ONE*,  
895 *9*(11). <https://doi.org/10.1371/journal.pone.0112644>

896 Huber, L., & Lenz, R. (1993). A test of the linear feature model of polymorphous concept discrimination with  
897 pigeons. *The Quarterly Journal of Experimental Psychology: Section B*, *46*(1), 1–18.

898 King, M. L., Groen, I. I. A., Steel, A., Kravitz, D. J., & Baker, C. I. (2019). Similarity judgments and cortical visual  
899 responses reflect different properties of object and scene categories in naturalistic images.  
900 *NeuroImage*, *197*(March), 368–382. <https://doi.org/10.1016/j.neuroimage.2019.04.079>

901 Kirkpatrick-Steger, K., Wasserman, E. A., & Biederman, I. (1998). Effects of Geon Deletion, Scrambling, and  
902 Movement on Picture Recognition in Pigeons. *Journal of Experimental Psychology: Animal Behavior*  
903 *Processes*, *24*(1), 34–46.

904 Larkey, L. B., & Markman, A. B. (2005). Processes of similarity judgment. *Cognitive Science*, *29*(6), 1061–1076.  
905 [https://doi.org/10.1207/s15516709cog0000\\_30](https://doi.org/10.1207/s15516709cog0000_30)

906 Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta*  
907 *Psychologica*, *102*(2–3), 221–245. [https://doi.org/10.1016/S0001-6918\(98\)00052-3](https://doi.org/10.1016/S0001-6918(98)00052-3)

908 Lawson, R., & Jolicoeur, P. (2003). Recognition thresholds for plane-rotated pictures of familiar objects. *Acta*  
909 *Psychologica*, *112*(1), 17–41. [https://doi.org/10.1016/S0001-6918\(02\)00099-9](https://doi.org/10.1016/S0001-6918(02)00099-9)

910 Leech, R., Mareschal, D., & Cooper, R. P. (2007). Relations as transformations: Implications for analogical  
911 reasoning. *The Quarterly Journal of Experimental Psychology*, *60*(7), 897–908.  
912 <https://doi.org/10.1080/17470210701288599>

913 Luzardo, A., Alonso, E., & Mondragón, E. (2017). A Rescorla-Wagner drift-diffusion model of conditioning and  
914 timing. In *PLoS Computational Biology* (Vol. 13, Issue 11). <https://doi.org/10.1371/journal.pcbi.1005796>

915 Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity.  
916 *Journal of Memory and Language*, *32*(4), 517–535. <https://doi.org/10.1006/jmla.1993.1027>

917 Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and*  
918 *Cognition*, *24*(2), 235–249. <https://doi.org/10.3758/BF03200884>

919 Martin-Ordas, G., Berntsen, D., & Call, J. (2013). Memory for distant past events in chimpanzees and  
920 orangutans. *Current Biology*, *23*(15), 1438–1441. <https://doi.org/10.1016/j.cub.2013.06.017>

921 Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*(2), 254–  
922 278. <https://doi.org/10.1037/0033-295X.100.2.254>

923 Mendes, N., Rakoczy, H., & Call, J. (2011). Primates do not spontaneously use shape properties for object  
924 individuation: A competence or a performance problem? *Animal Cognition*, *14*(3), 407–414.  
925 <https://doi.org/10.1007/s10071-010-0375-0>

- 926 Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-  
927 similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in*  
928 *Psychology*, 4(MAR), 128. <https://doi.org/10.3389/fpsyg.2013.00128>
- 929 Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental*  
930 *Psychology: Learning, Memory, and Cognition*, 10(1), 104–114. [https://doi.org/10.1037/0278-  
931 \*7393.10.1.104\*](https://doi.org/10.1037/0278-7393.10.1.104)
- 932 Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction.  
933 *Psychological Review*, 97(2), 185–200. <https://doi.org/10.1037/0033-295X.97.2.185>
- 934 Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model.  
935 *Psychological Review*, 101(4), 587–607. <https://doi.org/10.1037/0033-295X.101.4.587>
- 936 Pearce, J. M. (2008). *Animal Learning and Cognition: An Introduction*. Taylor & Francis.
- 937 Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between  
938 human and nonhuman minds. *Behavioral and Brain Sciences*, 31(2), 109–178.  
939 <https://doi.org/10.1017/S0140525X08003543>
- 940 Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative,  
941 critical review. *Annual Review of Psychology*, 58, 97–118.  
942 <https://doi.org/10.1146/annurev.psych.58.110405.085555>
- 943 Poldrack, R. A. (2020). The physics of representation. *Synthese*, 199(1), 1307–1325.  
944 <https://doi.org/10.1007/s11229-020-02793-y>
- 945 Rasmussen, A., Zucca, R., Johansson, F., Jirnhed, D. A., & Hesselow, G. (2015). Purkinje cell activity during  
946 classical conditioning with different conditional stimuli explains central tenet of Rescorla-Wagner  
947 model. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45),  
948 14060–14065. <https://doi.org/10.1073/pnas.1516986112>
- 949 Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of  
950 reinforcement and nonreinforcement. In H. Black, A & W. F. Prokasy (Eds.), *Classical Conditioning II*  
951 *Current Research and Theory* (Vol. 21, Issue 6, pp. 64–99). Appleton-Century-Crofts.  
952 <https://doi.org/10.1101/gr.110528.110>
- 953 Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16(3), 371–  
954 416. [https://doi.org/10.1016/0010-0285\(84\)90014-8](https://doi.org/10.1016/0010-0285(84)90014-8)
- 955 Santos, L. R., Miller, C. T., & Hauser, M. D. (2003). Representing tools: How two non-human primate species  
956 distinguish between the functionally relevant and irrelevant features of a tool. *Animal Cognition*, 6(4),  
957 269–281. <https://doi.org/10.1007/s10071-003-0171-1>
- 958 Santos, L. R., Sulikowski, G. M., Spaepen, G. M., & Hauser, M. D. (2002). Object individuation using  
959 property/kind information in rhesus macaques (*Macaca mulatta*). *Cognition*, 83(3), 241–264.  
960 [https://doi.org/10.1016/S0010-0277\(02\)00006-9](https://doi.org/10.1016/S0010-0277(02)00006-9)
- 961 Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to  
962 distance in psychological space. *Psychometrika*, 22(4), 325–345. <https://doi.org/10.1007/BF02288967>

- 963 Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210(4468), 390–398.  
964 <https://doi.org/10.1126/science.210.4468.390>
- 965 Shepherd, J. (2018). Intelligent action guidance and the use of mixed representational formats. *Synthese*,  
966 198(517), 4143–4162. <https://doi.org/10.1007/s11229-018-1892-7>
- 967 Simons, J. S., Lee, A. C. H., Graham, K. S., Verfaellie, M., Koutstaal, W., Hodges, J. R., Schacter, D. L., &  
968 Budson, A. E. (2005). Failing to get the gist: Reduced false recognition of semantic associates in  
969 semantic dementia. *Neuropsychology*, 19(3), 353–361. <https://doi.org/10.1037/0894-4105.19.3.353>
- 970 Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category Learning in Rhesus Monkeys: A Study of the  
971 Shepard, Hovland, and Jenkins (1961) Tasks. *Journal of Experimental Psychology: General*, 133(3),  
972 398–414. <https://doi.org/10.1037/0096-3445.133.3.398>
- 973 Taylor, E. G., & Hummel, J. E. (2009). Finding similarity in a model of relational reasoning. *Cognitive Systems*  
974 *Research*, 10(3), 229–239. <https://doi.org/10.1016/j.cogsys.2008.09.004>
- 975 Theves, S., Fernandez, G., & Doeller, C. F. (2019). The hippocampus encodes distances in multidimensional  
976 feature space. *Current Biology*, 29(7), 1226–1231.e3. <https://doi.org/10.1016/j.cub.2019.02.035>
- 977 Tomasello, M., & Call, J. (2007). *Primate Cognition*. Oxford University Press.
- 978 Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states—The question is  
979 which ones and to what extent. *Trends in Cognitive Sciences*, 7(4), 153–156.  
980 [https://doi.org/10.1016/S1364-6613\(03\)00035-4](https://doi.org/10.1016/S1364-6613(03)00035-4)
- 981 Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Origin of cultural cognition. *Behavioral and*  
982 *Brain Sciences*, 28, 675–735.
- 983 Toussaint, G. T., Matthews, L., Campbell, M., & Brown, N. (2012). Measuring musical rhythm similarity:  
984 Transformation versus feature-based methods. *Journal of Interdisciplinary Music Studies*, 6(1), 23–53.  
985 <https://doi.org/10.4407/jims.2012.12.002>
- 986 Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84(2), 127–190. [https://doi.org/10.1037/0033-](https://doi.org/10.1037/0033-295X.84.4.327)  
987 [295X.84.4.327](https://doi.org/10.1037/0033-295X.84.4.327)
- 988 Vonk, J. (2003). Gorilla (*Gorilla gorilla gorilla*) and orangutan (*Pongo abelii*) understanding of first- and  
989 second-order relations. *Animal Cognition*, 6(2), 77–86. <https://doi.org/10.1007/s10071-003-0159-x>
- 990 Wagner, A. R., & Brandon, S. E. (2000). A componential theory of Pavlovian conditioning. In R. R. Mowrer & S.  
991 B. Klein (Eds.), *Handbook of Contemporary Learning Theories* (pp. 23–64). Lawrence Erlbaum  
992 Associates Publishers. <https://doi.org/10.4324/9781410600691>
- 993 White, J. B. (2008). Self-other similarity judgment asymmetries reverse for people to whom you want to be  
994 similar. *Journal of Experimental Social Psychology*, 44(1), 127–131.  
995 <https://doi.org/10.1016/j.jesp.2006.10.015>
- 996 Zentall, T. R., Wasserman, E. A., & Urquioli, P. J. (2014). Associative concept learning in animals. *Journal of the*  
997 *Experimental Analysis of Behavior*, 101(1), 130–151. <https://doi.org/10.1002/jeab.55>
- 998