

Similarity and structured representation in human and nonhuman apes

Carl J. Hodgetts^{a,b,*}, James O.E. Close^{c,d}, Ulrike Hahn^e

^a Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

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

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

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

^e Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK

ARTICLE INFO

Keywords:

Similarity
Object representation
Features
Cognitive evolution
Perceptual discrimination

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.

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, Smith, Wilkie, López, & Shafir, 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, Goldstone, & Gentner, 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).

A major debate within human psychology concerns the role of *relations* in similarity: a table, for example, is not just a collection of features (tabletop, legs, colour etc.) but these features arranged in a

particular way. Human visual representation, therefore, is argued to involve so-called ‘structured representations’, that is, representations that involve both features and the relations between them (Biederman, 1987; Hafri & Firestone, 2021; Hahn, Chater, & Richardson, 2003; Markman & Gentner, 1993). Indeed, it has been shown that when two objects share a perceptual feature (e.g., ‘red’), it contributes more to human similarity judgements when it appears in corresponding positions of a relational structure – referred to as a match-in-place (Goldstone, 1994a). For instance, if presented with an image of two people wearing coloured hats and shirts, the pair will be perceived to be more similar if both hats are red and both shirts blue, than if the hat and shirt colours are swapped for one person in the pair, despite the overall feature set remaining unchanged.

Attempts to formally measure similarity, both within cognitive psychology and machine learning, have thus started to move toward ways of calculating similarity over structured representations (Hahn et al., 2003; Markman & Gentner, 1993). However, many of the most popular models, particularly in the context of animal learning (Pearce, 1994; Rescorla & Wagner, 1972), still treat stimulus representation as a matter of decomposing stimuli into individual, task-relevant features, whether in a feature vector (Tversky, 1977), or a spatial representation that represents items as points in a multi-dimensional space (Shepard, 1957).

* Corresponding author at: Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK.

E-mail address: carl.hodgetts@rhul.ac.uk (C.J. Hodgetts).

<https://doi.org/10.1016/j.cognition.2023.105419>

Received 23 June 2022; Received in revised form 17 February 2023; Accepted 18 February 2023

Available online 25 April 2023

0010-0277/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

As we will outline below, the only way relational structure is encoded in these models is by treating relations themselves as ‘features’, in particular, as ‘conjunctive features’ (i.e., ‘red’ and ‘hat’ becomes ‘red+hat’).

While such ‘featural’ representations appear too simplistic for humans – at least in some contexts (Goldstone & Medin, 1994; Hahn et al., 2003; Markman & Gentner, 1993) – the widespread use of featural coding schemes in the animal learning literature would seem to reflect an implicit (or even explicit) assumption that these may be adequate for some, or even *all*, nonhuman species. This assumption might have been fuelled further by the finding that nonhuman species have difficulty recognising purely relational similarities (Blough, 2001; George, Ward-Robinson, & Pearce, 2001; Haun & Call, 2009). The fact that nonhuman primates do not seem to be able to deal with such purely relational similarities (e.g., “equal/unequal”, or “sameness”) is of course distinct from the question of whether relational information impacts similarity judgements more generally. Needless to say, if nonhuman primates do not include relational information in their object representations, then such information cannot impact their similarity judgements. The fact that they might, however, as has been argued in some studies (Hopkins & Washburn, 2002; Huber & Lenz, 1993; Kirkpatrick-Steger, Wasserman, & Biederman, 1998; Zentall, Wasserman, & Urciuoli, 2014), does not prejudice if and how this relational information impacts perceived similarity. Just showing sensitivity to relational structure does not tell us much (if anything) about the similarity *gradients* that stem from matches or mismatches. To understand this, structural models of similarity from the human literature must be tested systematically on animal behaviour. Given the close connection between similarity and theories of representation outlined above, such an examination should elucidate not only similarity, in particular the limitations of feature-based models of similarity, but may also provide valuable insight into the mental object representations of nonhuman species.

1.1. The limits of featural coding

To address this question, we examined the nature of perceptual similarity judgements in three great ape species: humans, chimpanzees, and gorillas. Our main question was whether perceived similarity in chimpanzees and gorillas are, like humans, sensitive to ‘structure’. To this end, we devised a simple stimulus set that allowed us to adjudicate between featural and structural models of similarity. Before outlining these stimuli and the model predictions in detail, we will first describe some of the fundamental issues facing featural models of similarity. This is best done with reference to a particular set of items, so we will use example items from our actual stimulus set. Each item in our stimulus set involves a pair of geometric figures, such as those shown in Fig. 1. Broadly, featural models of similarity treat each psychologically relevant aspect of an object as a single component, and the entire object is represented by all relevant components – whether these are represented as a feature set, a feature vector, or a point in a multi-dimensional feature space (Shepard, 1980; Tversky, 1977). To illustrate: the left stimulus in Fig. 1 (*item 1*) might be represented by the feature set {square, circle, white, black}.

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

Such an approach runs into trouble because it leads to a proliferation of ‘features’: all the basic elements *and* all their possible combinations must be retained to avoid the hyper-specificity that combinations would otherwise bring. To illustrate, if one were to only consider

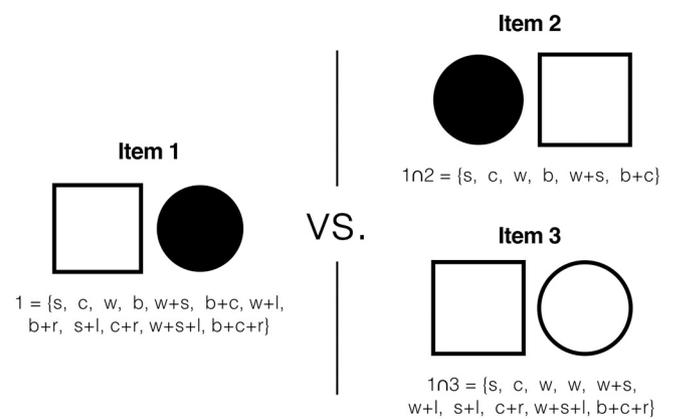


Fig. 1. A simple set of three stimuli (*items 1, 2 and 3*) for comparison. Below *item 1* is a featural description in terms of the basic features (s = ‘square’, c = ‘circle’, w = ‘white’, and b = ‘black’), the feature conjunctions ($w+s$ = ‘white+square’, $b+c$ = ‘black+circle’, etc.) and the features/conjunctions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

‘white+square+left’ as a single compound feature, then *item 1* and *item 2* no longer share any features and would thus appear maximally dissimilar. In other words, one needs to retain both the component features (‘white’) and the conjunctions (‘white+square’) to account for commonalities that appear across different positions and/or objects. This requirement not only leads to a combinatorial explosion (with the number of conjunctions determined by the binomial coefficient ‘ N choose k ’, which for 10 individual features will add a further 45 separate feature pairs, and at $N = 15$ a further 105) but all of these features potentially influence the similarity comparison and counting them in assessing ‘commonality’ itself can lead to counter-intuitive distortions. For example, such proliferations make it the case that *items 1* and *2* have considerably *fewer* features in common when compared to *items 1* and *3*.

Representation schemes that allow one to represent relations are known as structured representations (Gentner, 1983; for more recent literature, see e.g., Dourmas, Hummel, & Sandhofer, 2008; Dourmas & Martin, 2018; Poldrack, 2020; Shepherd, 2018). The most popular examples of such schemes are graph structures or multi-place predicates as found in first-order logic, for example, a 2-place predicate such as TO-THE-LEFT-OF(x,y). Turning relations into features means that a binary relation such as TO-THE-LEFT-OF(x,y) effectively becomes a feature such as TO-THE-LEFT-OF- $X(y)$. The crucial difference between these two schemes is that only the former, relational, representation separates out the relation and both its arguments in such a way that they can be accessed simultaneously and thus independently factored into the similarity comparison. TO-THE-LEFT-OF(x,y) might, for example, provide a relational match to TO-THE-LEFT-OF(q,r) thus providing a purely relational commonality across multiple object pairs such as those in Fig. 1. For the feature-based version of the same state of affairs, one is left simply with two distinct properties TO-THE-LEFT-OF- $X()$ and TO-THE-LEFT-OF- $Y()$. The same is ultimately true of the kinds of conjunctive coding schemes that are popular in associative and connectionist models (Blumberg & Sokoloff, 2001; Dickinson, 2012; Gluck & Bower, 1988).

It is for this reason that many theories concerned with the representation of real-world objects or events – whether these are faces, scenes, sentences, or extended narratives – assume that these cannot be represented on purely featural schemes (Biederman, 1987; Hahn et al.,

Table 1

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

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

2003; Markman & Gentner, 1993). Instead, they seem to require structured representations: complex representations of objects, their parts and properties, and – crucially – the interrelationships between them, that cannot be boiled down to either lists of features or points in space.

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

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

Though difficult, we seek to show that such evidence against conjunctive coding is possible. The strategy for doing so is already hinted at in Fig. 1. Specifically, the contrast between featural and structural representation can be pursued successfully over a suitable set of items. While ‘featural’ reconstructions may seem *locally plausible* when considering just one or two comparisons, they can be shown to be *globally implausible* over an entire set of items. The crucial ‘trick’ here is to take whatever conjunctive features ‘do the work’ in one comparison and then identify another contrast that can be added to the set of

similarity comparisons under consideration where those conjunctive features lead to difficulties, generating implausible predictions of similarity. The stimulus set presented in this paper was designed to do just that.

1.2. Devising a stimulus set for contrasting featural and structural models

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

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

The ability to also code conjunctive features, such as ‘blue+square’, is a strategy designed to address this issue (see e.g., Wagner & Brandon, 2000, and references therein). Like the FEAT model described above, this conjunctive feature model (C-FEAT) can form multiple matches, such that basic features and conjunctions in one stimulus can match with multiple features/conjunctions in the other. While such an approach is

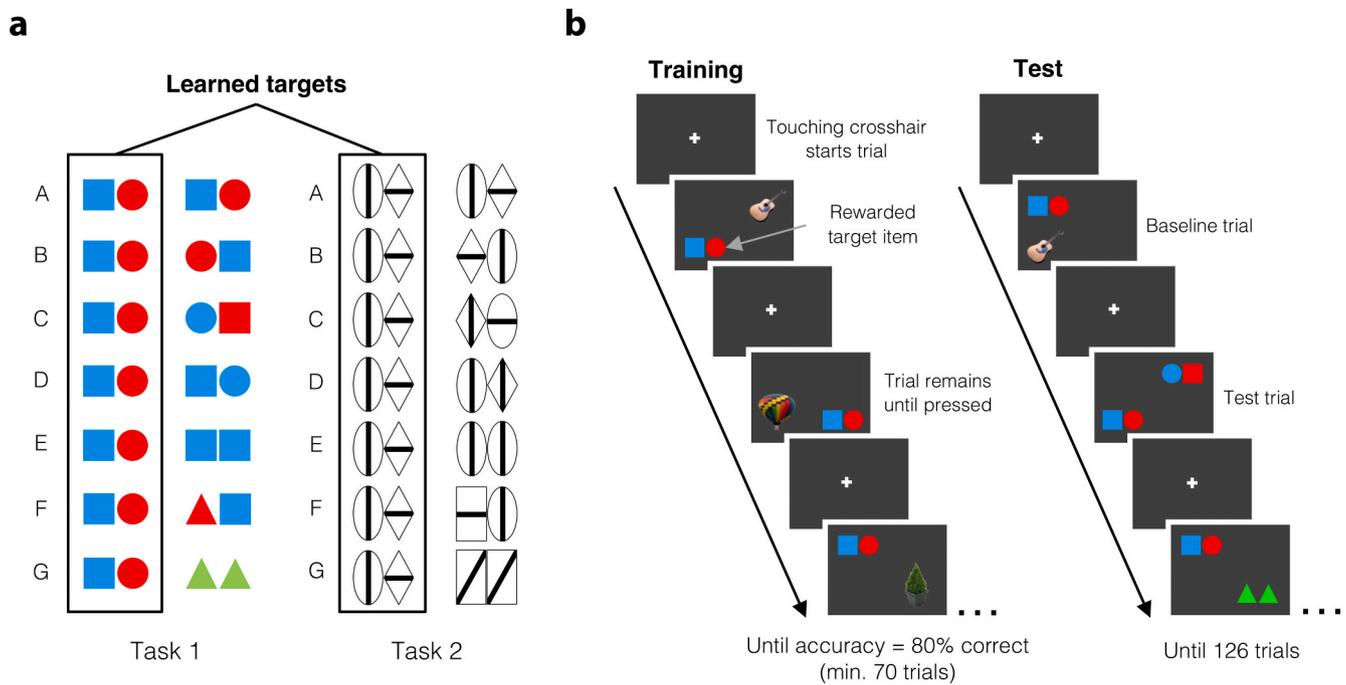


Fig. 2. (A) The stimulus comparisons used in Task 1 (left) and Task 2 (right). Each task has a target item comprising features A and B on dimension 1 (Task 1 = colour [blue/red]; Task 2 = line orientation [vertical/horizontal]) and A and B on dimension 2 (Task 1 = shape [square/circle]; Task 2 = shape [oval/diamond]). Alternative combinations of these features make up the comparison stimuli, as outlined in the Methods. (B) Trial schematic for the nonhuman ape training session is shown on the left. This phase established their preference for the target items shown in panel A. Subjects received a food reward (grape) for selecting the target item over the everyday object stimulus. Training terminated when subjects reached the required criterion (80% correct). In the main experiment, baseline trials (target item vs. everyday object) were intermixed with ‘test trials’. For test trials, targets items were paired with one of seven stimuli from the set (which vary in their similarity to the target). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sufficient for distinguishing comparisons where features have swapped across objects (i.e., comparison C vs. comparison B in Table 1), such models cannot distinguish between identical items (comparison A) and a ‘swap’ (comparison B), unless they also code for relative spatial location of features. Clearly the relative spatial location of features is relevant – at least for humans (Hodgetts et al., 2009). To capture this, we can specify a new conjunctive *spatial* feature model (CS-FEAT), in which relative spatial features are also added, both as single features but also as components of more extended *conjunctive* features (e.g., ‘left’, ‘blue+left’, and ‘blue+square+left’ get added to our feature set or vector; e.g., see George et al., 2001; George & Pearce, 2003). This new model (CS-FEAT) now distinguishes comparisons A and B (i.e., identity and swap) at the expense of over-matching ‘down the road’ (see Table 1). Namely, the predicted similarities of comparisons D and E are now greater than B. Limiting feature matches to certain spatial locations (i.e., ‘blue’ matches ‘blue’ if and only if blue is in a corresponding spatial position) via a spatial-feature model (S-FEAT) seems plausible to address this, but renders equally *dissimilar* comparisons B, F, and even G! These stimulus materials thus draw out the fundamental limitation of features in capturing relational information. Specifically, tweaking the feature set to include conjunctive and/or relational features may seem *locally plausible* when considering just one or two comparisons, but can be shown to be *globally implausible* over an entire set of items.

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

‘red’) form parts of whole objects (the hat), which in turn play a specific role within the broader relational structure. The alignment process itself (see e.g., Falkenhainer, Forbus, & Gentner, 1989; Gentner, 1983; Gentner & Markman, 1997) seeks to form matches that are structurally consistent across two representations, which requires that: a) an element in one representation must match with at most one element in the other representation (*one-to-one mapping*); and b) wherever relations are placed in correspondence, their respective arguments are also placed in correspondence (*parallel connectivity*) (Falkenhainer et al., 1989).

Underpinning these structural alignment models, as demonstrated in the ‘hat’ and ‘coat’ example above, is the classic distinction between matches for elements that have been placed in correspondence (matches-in-place, or MIPs) and matches for elements that do not correspond (matches-out-of-place, or MOPs). Given the established impact of MIPs on similarity ratings in humans (Markman & Gentner, 1996), including within the stimulus domain used here (Hodgetts et al., 2009), our alignment model was based on the number of MIPs. Consistent with the *one-to-one mapping* constraint, a feature match on our MIP model only “counts” if the objects themselves have been placed in correspondence.¹ For comparison C, for instance, the lower-level features ‘blue’ and ‘square’ (which make up the left-hand *object* in the

¹ 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 adopted a MIP-based implementation of structural alignment in this paper.

target stimulus) map on to two separate *objects* in the right stimulus. Given the *one-to-one mapping* constraint, such many-to-one mappings, where a single feature/object in one stimulus is matched with two or more features/objects in a second stimulus, are not permitted under this MIP model. As a result, only a single match is counted – e.g., the blue square is matched with the blue circle on the basis of colour, and likewise the red circle is matched with only one object in the right stimulus: the ‘red square’ based on the shared property ‘red’. This adherence to one-to-one mappings at the level of objects allows the MIP model to distinguish comparisons B and C, and also does not lead to the profound proliferation of features for other comparisons in the set.

The transformational model of similarity, or ‘Representational Distortion’ (RD) (Hahn et al., 2003), proposes that perceived similarity emerges from the complexity, or ‘effort’, of transforming the mental representation of one object or event into another. Measures of transformational complexity, and thus (dis)similarity, may range from continuous spatial transformations (e.g., translation, rotation, etc), as seen in models of visual object recognition (Graf, 2006; Hahn, Close, & Graf, 2009; Lawson, 1999; Lawson & Jolicoeur, 2003), to sets of simple operations (insert, delete, swap, etc), which can then be combined into longer codes to capture more complex transformational relationships (Hahn et al., 2003; Hahn & Bailey, 2005; Hodgetts et al., 2009). Conceptually, a single transformation may act upon individual features (or indeed continuous feature dimensions), whole ensembles of features, or manipulate the interrelationship between features or objects (i.e., structure; Hahn et al., 2003). In past empirical work, it has been shown that transformational model predictions can capture accurately human similarity ratings (Hahn et al., 2003; Hodgetts et al., 2009), speeded same-different judgements (Hodgetts & Hahn, 2012), and even analogical reasoning (Leech, Mareschal, & Cooper, 2007). Critically, it has also been shown that transformations provide superior fits of human similarity data when compared to basic feature models (Hahn et al., 2003; Toussaint, Matthews, Campbell, & Brown, 2012) and models of structural alignment (Hodgetts et al., 2009; see also Larkey & Markman, 2005).

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

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

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

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

1.3. The current investigation

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

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

2. Methods

2.1. Participants

Five chimpanzees (*Pan troglodytes*; 1 male, 4 female) and three gorillas (*Gorilla gorilla*; 3 females) took part in Task 1 (8 apes in total). One chimpanzee (Trudi) did not complete Task 2 as she did not reach criterion (80% correct) during the learning phase, resulting in $N = 4$ chimpanzees in Task 2 (7 apes in total). All subjects were housed at the Wolfgang Köhler Primate Research Center at Zoo Leipzig (Germany). They lived in social groups with conspecifics and had access to indoor and outdoor areas designed to be appropriate for their species. All subjects had touchscreen experience. Apes were tested individually within a familiar, indoor room, with the exception of one gorilla

(Viringika). In Viringika's case, her young daughter accompanied her in the testing area; no obvious disruption to Viringika's performance was noted from this arrangement. Ten human volunteers (3 males, 7 females; mean age = 31 years; SD = 6.93) were tested at Cardiff University, School of Psychology. This was undertaken with the understanding and written consent of each participant. All participants had normal or corrected to normal vision. Informed consent was obtained after the nature and possible consequences of the study were explained to participants, in accordance with the local research ethics committee at Cardiff University. Further, the animals' care was in accordance with institutional guidelines at the Max Planck Institute for Evolutionary Anthropology, and Zoo Leipzig.

2.2. Stimulus design

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

2.3. Apparatus

The nonhuman tasks were presented on a 15-in. LCD monitor, mounted in a custom-built metal holder. The monitor was situated behind a 15-in. infrared touchscreen frame. The touchscreen frame replaced one of the standard safety panels that separated the nonhuman apes from the experimenter, becoming the de facto safety panel. Situated in front of the touchscreen frame was a thin Plexiglas panel with five circular finger holes cut into it. These were in the centre, top-left, top-right, bottom-left, and bottom-right corners of the panel. As well as forming an additional safety panel, the Plexiglas panel also enabled safe touching of the touchscreen frame by the apes. Located on the floor to either side of the touchscreen-metal frame, and facing toward the subjects, were two small speakers used to present audio feedback. A PC, outputting a display resolution of 1280×1024 pixels, was connected to the LCD monitor and touchscreen frame, and E-Prime 2.0 (Psychology Software Tools, Inc., Sharpsburg, PA) was used to run the experiment. Food rewards were fed by hand when a correct response was made through a plastic tube located next to the touchscreen-metal frame. The setup was identical for the chimpanzees and the gorillas. The human version of both tasks was presented on a 13-in. laptop outputting a display resolution of 1280×1024 pixels, and responses were made using a mouse.

2.4. Non-human tasks

2.4.1. Task 1 - training phase

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

2.4.2. Task 1 - test phase

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

2.4.3. Task 2 - training phase

The same procedure as Task 1 was used with the following exceptions: 1) a trial began with a black fixation point (150×150 pixels); 2) the background colour of the screen was always white; and 3) the screen positioning of the target stimulus and real-world object was pseudo-random, with the constraint that the target stimulus could not be

presented on the left or right of the real-world object more than two times in a row. While there was no evidence of the nonhuman apes developing a side bias in Task 1, this control was added to reduce the possibility of a side bias being developed in Task 2.

2.4.4. Task 2 - test phase

The same procedure as Task 1 was used, with the exceptions noted above and one further exception: only a maximum of four (unrewarded) test trials could be presented in a row. The reason that this was implemented was because it was possible, theoretically, for up to 14 unrewarded test trials to appear consecutively in Task 1 (though this never occurred). Based on this change to the experimental design, one block in Task 2 consisted of seven test trials – as in Task 1 – and between nine and 12 non-test trials (i.e., target shape versus everyday object picture). In each session, six blocks were completed. Overall, each session consisted of 126 trials (as in Task 1): 84 baseline trials (including the initial 21 trials) and 42 test trials.

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

2.5. Human tasks

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

3. Results

3.1. Structure matters for human and nonhuman species

We first examined the relationship between perceived similarity in each ape species and the predictions of our featural and structural similarity models. Assuming a power law relationship between similarity and response time (Cohen & Nosofsky, 2000; Hodgetts & Hahn, 2012), the best fitting similarity model for human participants is a structural model, RD, across both tasks; this model captures 98% of the variance for Task 1, and 88% of the variance in Task 2 (Table 2). For the nonhuman apes, the best fitting model is again RD for Task 1, capturing 58% of the variance, but for Task 2 it is outperformed by the CS-FEAT model, with 98% of the variance (the fits for each tested model can be found in Table 2).

This suggests that chimpanzees and gorillas processed the Task 2 materials, with their more artificial, shape-related dimensions (outer shape/inner line orientation), in a fundamentally different way to the more ecologically familiar shape-colour combination found in Task 1. In other words, there is a task difference on the basis of materials for chimpanzees and gorillas. While their performance is aligned to humans on one task, showing evidence for structure, their performance looks “featural” on Task 2. In the remainder, we provide analyses that follow up on this point, while further examining the putative role of structure in the perceived similarity of nonhuman apes.

Table 2

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

Default model	Task 1				Task 2			
	Human		NHA		Human		NHA	
	Linear R^2	Power R^2						
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	0.98*	0.78
S-FEAT	0.32	0.22	0.23	0.21	0.17	0.11	0.81	0.80*
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	0.73*	0.98*	0.58*	0.80*	0.85*	0.88*	0.57	0.57

Weighted model	Human		NHA		Human		NHA	
	Linear R^2	Power R^2						
	FEAT	0.36	0.43	0.23	0.20	0.42	0.47	0.74
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	0.97*	0.88*
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	0.75*	0.92*	0.70*	0.87*	0.85*	0.88*	0.79	0.78

3.2. Task differences emerge in nonhuman apes but not humans

The difference between these tasks, as well as an across-species structure sensitivity in Task 1, is confirmed by a global analysis that combines all the goodness of fit values (R-squared) for the ‘featural’ models found in Table 1 into a single feature measure, and likewise all ‘structural’ models and their goodness of fit into a structure measure. To derive a simple metric of ‘structurality’ from this model space, we calculated a difference score between the structure and feature measures for each individual subject. A positive score on this metric indicates that the pattern of perceived similarity across comparisons – at an individual-level – is better captured by structural models of similarity.

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

When this metric was evaluated for Task 2 – in which human and nonhuman apes were weakly correlated in terms of general patterns of similarity (Fig. 3C; Supplementary Results) – the apes appeared more featural overall (Fig. 3A). The humans, however, were numerically even more structural in Task 2 (Task 1 mean = 0.26; Task 2 mean = 0.29). These observations were confirmed statistically, such that the structurality difference between human and nonhuman ape species was strongly dependent on task stimuli ($F(1, 15) = 14.99, p = 0.002, \eta_p^2 = 0.5, BF_{10} = 2793.9$; Fig. 3A). Follow-up Welch *t*-tests revealed no significant

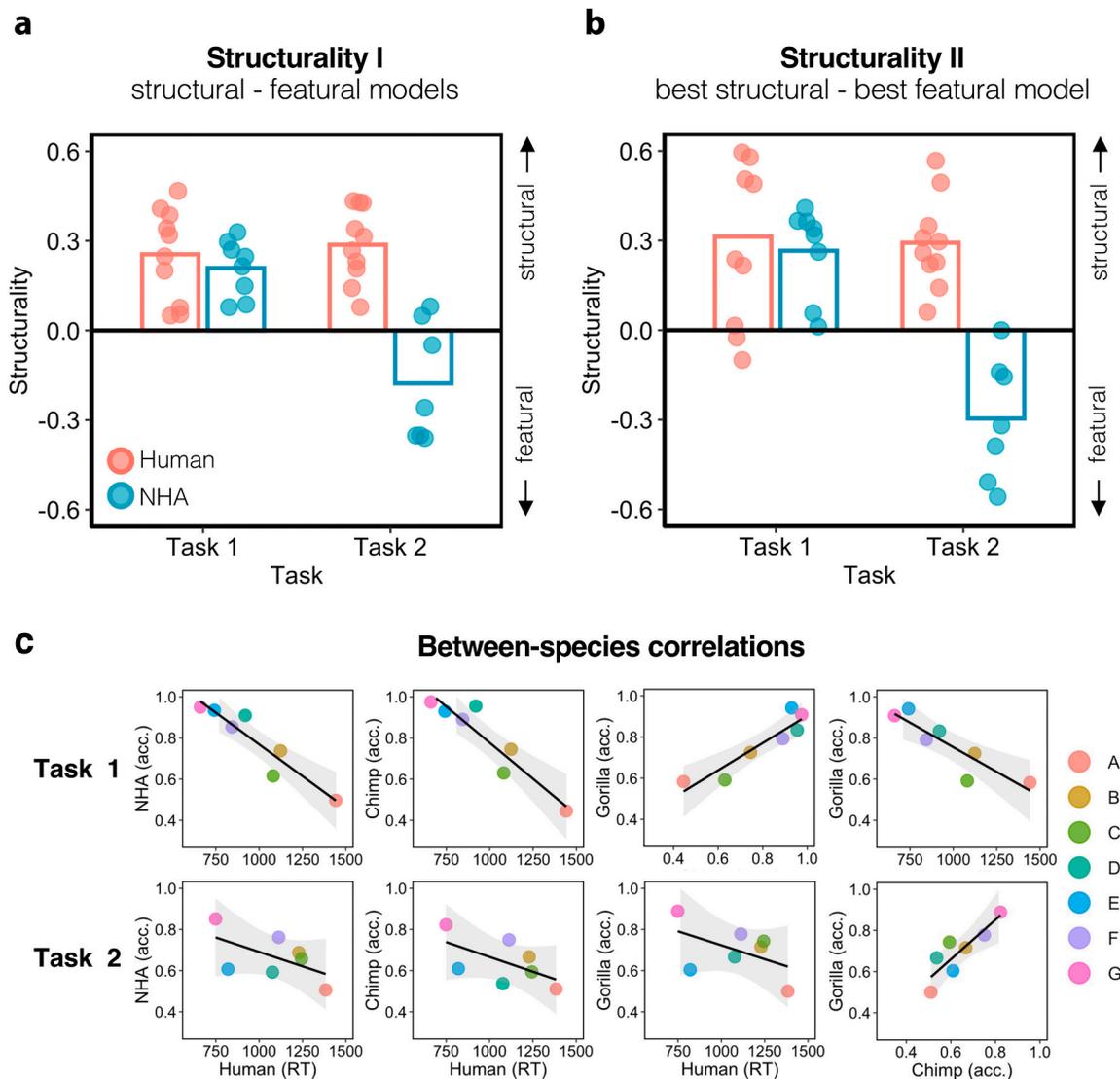


Fig. 3. Sensitivity to structure in both human and nonhuman ape species (labelled 'NHA' in the figure). (A) 'Structurality' scores were calculated by fitting each model to similarity data in each individual subject (response time (RT) for humans and accuracy (acc.) for nonhumans) and subtracting the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

difference in the structurality measure in Task 1 ($p = 0.45$; $BF_{10} = 0.55$) but a significant difference in Task 2 ($p < 0.001$; $BF_{10} = 594.6$). Furthermore, while the human structurality scores did not differ between tasks ($p = 0.62$; $BF_{10} = 0.35$), the nonhuman ape group were significantly more 'featural' in Task 2 ($p = 0.005$; $BF_{10} = 10.82$). The same interaction emerges when considering the single best-fitting similarity model for each subject (shown in Table S1). Here, the best fitting model is structural for the majority of the human participants in both Task 1 (8/10) and Task 2 (10/10), and for all chimpanzees and gorillas in Task 1 (8/8). The featural models, however, provide the best fit for 5/7 nonhuman apes in Task 2 (Table S1).

As the structurality measure includes a larger pool of featural models, it may underestimate the role of features by also considering those which provide a poor fit of the data in a given subject. To address this possibility, we recalculated the structurality measure in each subject (across species) and subtracted the single best fitting feature model from

the best fitting structural model (see Fig. 3B). The same interaction was observed ($F(1, 15) = 18.17$, $p = 0.001$, $\eta_p^2 = 0.46$, $BF_{10} = 1821.27$; Fig. 3B), such that no significant difference was seen between groups in Task 1 ($p = 0.65$; $BF_{10} = 0.47$), but a difference was found in Task 2 ($p < 0.001$; $BF_{10} = 2129.69$).

3.3. Apes are sensitive to spatial location

The same picture also emerges when we consider individual item comparisons. As noted in the Introduction, the key challenge for featural models is distinguishing between identity and swap, that is, stimulus comparisons A and B (Table 1) – a distinction that can readily be captured by structural models. In a featural model, however, this can only be achieved by adding some sensitivity to spatial location, such as the relative position of objects within the stimulus (see Section 1.2). But this then invariably creates problems with comparisons E, F and G in

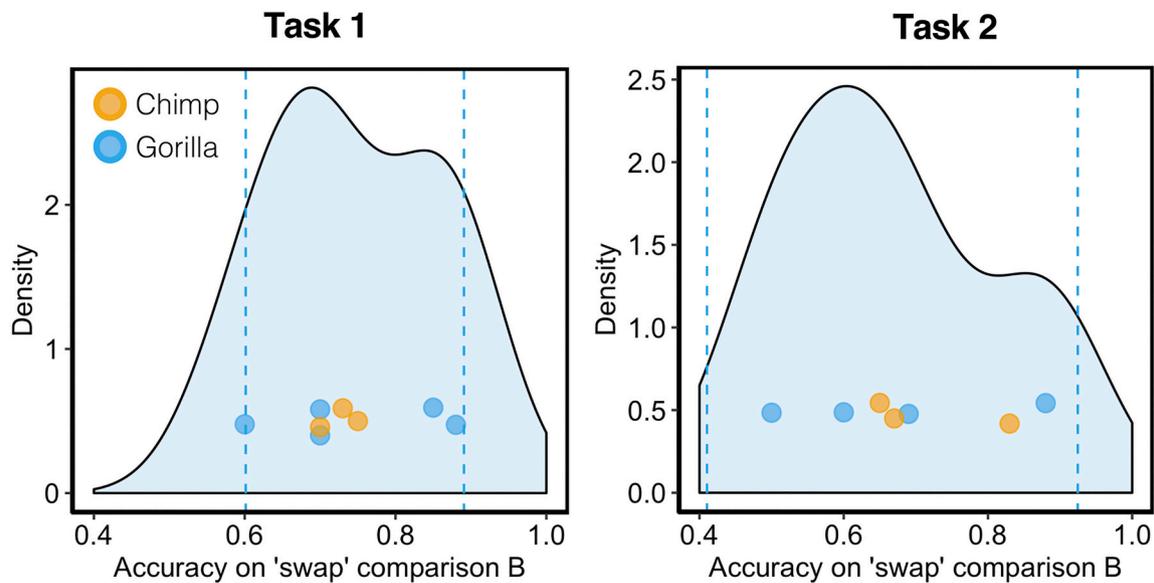


Fig. 4. Gorillas and chimpanzees are similarly sensitive to structure. The accuracy for each nonhuman subject on the 'swap' comparison *B* is shown for (A) Task 1 and (B) Task 2. The gorillas are shown to be within the chimpanzee distribution on both tasks. The density distribution and 95% confidence interval for the chimpanzees is shown in blue. The individual data points for the chimpanzees and gorillas are depicted by blue and orange markers, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1, where the specificity introduced by spatial features (to distinguish identity and swap) distorts the similarity across the set as a whole. Notably, all the species groups can distinguish between identity and swap (i.e., comparisons *A* and *B*; Fig. S3). The nonhuman apes, for instance, are exactly at chance for comparison *A* but are significantly more accurate at identifying the target item for the 'swap' comparison in both Task 1 ($p = 0.002$; $BF_{10} = 29.6$) and Task 2 ($p = 0.007$, $BF_{10} = 9.3$). Incidentally, though it has been claimed that gorillas are less sensitive to structure than chimpanzees (Haun & Call, 2009), on our items all three gorillas were able to distinguish identity and swap above chance and indeed fall within the centre of the chimpanzee distribution on both tasks (Fig. 4).

For humans, the pair of stimuli in comparison *E* are highly dissimilar in both tasks (Fig. S3), but the feature models with spatial sensitivity (see 'S-FEAT' and 'CS-FEAT' in Table 1) fail to capture this by rendering the items in *E* to be more similar than those in *B* (e.g., the swap). This limitation of the feature models, however, is relevant to performance in the nonhuman group; while the stimuli in comparison *E* are perceived to be dissimilar in Task 1, 3/7 nonhuman apes found the items in *E* to be as similar as those in identical comparison *A*! This indicates difficulty in accurately encoding relational information for stimuli based purely on line information, giving rise to the interactions observed in Fig. 3. Note, however, that these subjects' behavioural responses to *E* do not reflect a simple 'blindness' to one of the two feature dimensions in Task 2 (i.e., shape or line orientation), as we detail next.

3.4. Species-specific differences in dimensional salience

The relevance of specific feature dimensions to perceived similarity can be assessed by deriving model predictions for each dimension separately and then giving them different weight in determining similarity (100% colour, 90% colour/10% shape, and so on; see Hodgetts et al., 2009). This allows us to examine the performance of models for dimension 1 only, dimension 2 only, and all relative weightings in between (see Table 2 and Fig. S1A). Exploration of the best-fitting dimension weights for the well-performing models suggests that humans are sensitive to both colour and shape in Task 1, whereas chimpanzees and gorillas are, if anything, slightly more sensitive to

colour – an observation that resonates with previous findings (Mendes et al., 2011). Importantly, the same analyses for Task 2 suggest that, in contrast to human subjects, the nonhuman apes struggled to incorporate the orientation of the inner line for the Task 2 stimuli (Fig. S1A). This finding is also confirmed in a model-independent analysis that examines patterns of equivalence between comparisons that would be obtained if participants were blind to one or both of the dimensions (see Supplementary Results and Fig. S1B).

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

3.5. Common and distinctive features

Finally, it is worth commenting on the performance of Tversky's Contrast Model (Tversky, 1977). While virtually unknown in the animal literature, it is the most well-known featural model in human similarity research. The crucial feature of the Contrast Model is that similarity is a function of both the common and the unique features of each object in a comparison. Unlike the featural models outlined in Table 1, therefore, it does not just look for commonalities but also factors in those feature or attributes that distinguish objects (see Supplementary Methods for a full description of the model, and Tables 1 and S2 for overall and individual-level fits, respectively). Given the nature of our stimuli, the difference between counting just matches, and counting matches and mismatches is more apparent than real: the key value of Tversky's Contrast Model is

realised in situations where object descriptions differ in complexity such that the size (and weight) of the sets of distinctive features of each object vary. This, for example, is what allows the model to capture asymmetries in the context of directional similarity judgements. But these conditions are not met in our stimulus materials: every stimulus item consists of exactly two components, each of which has a colour and a shape. The number of matches is thus taken from the same wider set of possibilities in each case, so that enumerating the matching features effectively identifies the complement set of the non-matching features as well. As a result, we would not necessarily expect this model to perform vastly better than other featural models for our stimulus domain, even if contrasting features were relevant to all our participants. Indeed, while the Contrast Model is the best performing featural model for the humans on both Task 1 and Task 2, it is not the best featural model for nonhuman subjects (see CMOD-SM in Table 1). In fact, no single featural model performs best across Task 1 and Task 2 for the nonhuman apes; a finding that obtains both overall *and* at the individual level (see Tables S1–2). Similar to the individual-level fits reported in Section 3.2, the best fitting model is structural for the majority of human participants in Task 1 (7/10) and Task 2 (10/10) when the Contrast Model is also considered (Table S2). For nonhuman apes, the Contrast Model does not provide the best fit in any single subject for Task 1 or Task 2 (Table S2). Finally, the key interaction between group and task in Fig. 3 is retained when the Contrast Model is included in the structurality scores ($F(1, 15) = 12.8, p = 0.003, \eta_p^2 = 0.46; BF = 1164.935$).

4. Discussion

Contemporary models of human similarity place critical emphasis on the role of ‘structure’ in the representation of real-world objects and events (Gentner & Markman, 1997; Hahn et al., 2003), meaning that both individual elements and, critically, how they are interrelated can influence perceived similarity. Despite the relevance of structural information to human perceptual similarity, object recognition, analogical reasoning, and so on, dominant models of animal learning and cognition (e.g., Pearce, 1994; Rescorla & Wagner, 1972) still assume that nonhuman similarity and generalization can be readily captured by simple, feature-based representation schemes. Challenging this view, we have systematically demonstrated in this study that structural models of similarity can precisely capture, like in adult humans, patterns of similarity judgements in nonhuman species, specifically great apes (chimpanzees and gorillas). By applying a novel measure of ‘structurality’ – derived by contrasting a set of featural models with a set of established structural models of similarity – we found that all nonhuman apes demonstrated sensitivity to structure when discriminating basic object stimuli (coloured shapes). This provides strong evidence that nonhuman ape species utilise complex object representations that go beyond basic feature sets.

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

While all nonhuman apes demonstrated a striking sensitivity to structure for coloured shapes, they were notably more *featural* in Task 2,

which involved line information only (inner and outer contours). Additional model-based and model-free analyses suggested that nonhuman apes had particular difficulty integrating the outer shape with the inner line orientation. This interesting result aligns with prior work that highlights a central role for colour (versus shape) information in nonhuman primate object individuation and discrimination (Mendes et al., 2011; Santos, Sulkowski, Spaepen, & Hauser, 2002). While past cross-species work has found some congruence with human results with respect to the feature binding of colour and shape (e.g., Buračas & Albright, 1999; Cook, 1992),² research with other types of ‘features’ (such as the distinction between local and global features) has suggested the possibility of significant cross-species differences (e.g., Fagot & Tomonaga, 1999; Hopkins & Washburn, 2002) in how attributes of compound stimuli are perceived. In the context of our central question about the representation of relational structure, the contrast between performance in Task 1 and Task 2 highlights how the functional relevance of different forms of lower-level stimulus information may impact on the expression of higher-order cognitive processes. For example, while colour was highly relevant in this visual discrimination paradigm, it may be that shape information would be weighted more in contexts where it is ecologically/functionally relevant, such as when selecting stimuli for action, as seen in tool selection (Santos, Miller, & Hauser, 2003). Importantly, while the nonhuman apes failed to show structural sensitivity to the line-only stimuli of Task 2, they did show evidence of attending to both ‘dimensions’ (outline shape/inner line) in their discrimination. This suggests that the difference between the two tasks may reflect a performance difficulty rather than an in-principle limitation. Specifically, the highly salient, accessible, and familiar dimension of colour may be easier to process, leaving greater resources for the integration of the second dimension (even human participants were somewhat slower in Task 2, hinting at a visual processing advantage for the coloured objects in Task 1). In other words, the task differences for the nonhuman apes may reflect the kinds of performance differences seen widely for humans in cognitive and developmental psychology across familiar and less familiar materials. It thus seems possible that greater sensitivity to structure in Task 2 would emerge after more extensive training with the basic discrimination. This dissociation suggests also that requiring the binding mechanism to be wholly independent of the elements to-be-bound (as assumed to be definitional of structured representation in Dumas & Martin, 2018, p. 169) may be too strong a constraint for the meaning of structure in nonhuman species. Indeed, our results might be taken to suggest that it would be empirically and conceptually fruitful to not take the issue of structure in nonhuman primates as all or none (Dumas & Martin, 2018).

That said, the results from Task 1 do mean that models that do not include relational information are too restrictive for nonhuman apes, at least some of the time. Featural or vector-based models of stimulus representation, and as a consequence similarity, continue to dominate cognitive psychology (Ashby, 1992; Galesic, Walkyria Goode, Wallsten, & Norman, 2018; Hout, Goldinger, & Brady, 2014), animal learning (Hall & Rodríguez, 2017; Holmes, Chan, & Westbrook, 2019; Luzardo, Alonso, & Mondragón, 2017; Pearce, 2008; Rasmussen, Zucca, Johansson, Jirenhed, & Hesselow, 2015), cognitive neuroscience (King, Groen, Steel, Kravitz, & Baker, 2019; Mur et al., 2013; Theves, Fernandez, & Doeller, 2019) and machine learning (Hamel, 2009). Though arguments about the empirical adequacy of such representation schemes have been repeatedly made, they continue to be popular, arguably, because it is empirically difficult to provide compelling evidence for the role of structure in determining visual similarity. For one, this stems from the difficulty of excluding ‘mimicry’ of seeming sensitivity to structure through featural models. However, it stems also from the idea that

² 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, Minda, & Washburn, 2004).

structural similarities may be restricted in their role to ‘higher-level’ contexts, such as those involving some form of analogical reasoning (Hahn, 2014). This has made it seem plausible that structure does not matter to visual similarity in most, or all, nonhuman species. The present results refute this. Relational information has often figured as a fault line between researchers who view nonhuman species as capable of high-level abilities, such as causal reasoning (Beckers, Miller, De Houwer, & Urushihara, 2006; Blaisdell, Sawa, Leising, & Waldmann, 2006; Call, 2004), theory of mind (Tomasello, Call, & Hare, 2003), imitation (Tomasello, Carpenter, Call, Behne, & Moll, 2005) or mental time travel (Clayton & Dickinson, 2009; Martin-Ordas, Berntsen, & Call, 2013), on the one hand, and those who view seeming demonstrations of such abilities as the product of simpler associative processes (Dwyer, Starns, & Honey, 2009; Heyes, 1998, 2001; Penn & Povinelli, 2007). In this context, the present results emphasise the need to distinguish carefully between relations embedded within object and/or event representations and true higher-order relational reasoning (Penn et al., 2008). Indeed, these results emphasise the need to develop more realistic models of the former, which include relational structure, even in species that show no evidence of the latter.

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

Funding sources

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

CRediT authorship contribution statement

Carl J. Hodgetts: Conceptualization, Investigation, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing. **James O.E. Close:** Conceptualization, Investigation, Formal analysis, Writing – review & editing. **Ulrike Hahn:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing.

Data availability

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

Acknowledgements

We would like to thank John Pearce, John Aggleton, Craig Hedge, Petroc Sumner and the Cardiff University School of Psychology Cognitive Group for their helpful discussion and valuable feedback. In addition, we appreciate the comments of Daniel Povinelli on an earlier version of this manuscript. Finally, we would like to thank Josep Call and the keepers of Zoo Leipzig for their support and assistance in this research.

Appendix A. Supplementary data

The Supplementary Material contains supplementary methods (additional model descriptions), supplementary results (between-species correlations and model-free analyses of dimensional sensitivity),

and supplementary figures/tables.

References

- Ashby, F. G. (1992). In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition*. Psychology Press.
- Beckers, T., Miller, R. R., De Houwer, J., & Urushihara, K. (2006). Reasoning rats: Forward blocking in pavlovian animal conditioning is sensitive to constraints of causal inference. *Journal of Experimental Psychology: General*, 135(1), 92–102. <https://doi.org/10.1037/0096-3445.135.1.92>
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115–117. <https://doi.org/10.1037/0033-295X.94.2.115>
- Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, 311(5763), 1020–1022. <https://doi.org/10.1126/science.1121872>
- Blough, D. (2001). The perception of similarity. In R. G. Cook (Ed.), *Avian Visual Cognition*. Comparative Cognition Press.
- Blumberg, M. S., & Sokoloff, G. (2001). Do infant rats cry? *Psychological Review*, 108(1), 83–95. <https://doi.org/10.1037//0033-295X>
- Buracas, G. T., & Albright, T. D. (1999). Covert visual search: A comparison of performance by humans and macaques (*Macaca mulatta*). *Behavioral Neuroscience*, 113(3), 451–464. <https://doi.org/10.1037/0735-7044.113.3.451>
- Call, J. (2004). Inferences about the location of food in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, and *Pongo pygmaeus*). *Journal of Comparative Psychology*, 118(2), 232–241. <https://doi.org/10.1037/0735-7036.118.2.232>
- Cherries, E. W., Newman, G. E., Santos, L. R., & Scholl, B. J. (2006). Units of visual individuation in rhesus macaques: Objects or unbound features? *Perception*, 35(8), 1057–1071. <https://doi.org/10.1068/p5551>
- Clayton, N. S., & Dickinson, A. (2009). Mental time travel: Can animals recall the past and plan for the future?. In 4(August). *Encyclopedia of animal behavior* (pp. 438–442). <https://doi.org/10.1016/B978-0-08-045337-8.00094-2>
- Cohen, A. L., & Nofosofsky, R. M. (2000). An exemplar-retrieval model of speeded same-different judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1549–1569. <https://doi.org/10.1037/0096-1523.26.5.1549>
- Cook, R. G. (1992). Dimensional organization and texture discrimination in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 18(4), 354–363. <https://doi.org/10.1037/0097-7403.18.4.354>
- Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 367(1603), 2733–2742. <https://doi.org/10.1098/rstb.2012.0220>
- Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, 115(1), 1–43. <https://doi.org/10.1037/0033-295X.115.1.1>
- Doumas, L. A. A., & Martin, A. E. (2018). Learning structured representations from experience. In (1st ed.), 69. *Psychology of learning and motivation—Advances in research and theory*. Elsevier Inc. <https://doi.org/10.1016/bs.plm.2018.10.002>
- Dwyer, D. M., Starns, J., & Honey, R. C. (2009). “Causal reasoning” in rats: A reappraisal. *Journal of Experimental Psychology: Animal Behavior Processes*, 35(4), 578–586. <https://doi.org/10.1037/a0015007>
- Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, 21(4), 449–498. <https://doi.org/10.1017/S0140525X98001253>
- Fagot, J., & Parron, C. (2010). Relational matching in baboons (*Papio papio*) with reduced grouping requirements. *Journal of Experimental Psychology: Animal Behavior Processes*, 36(2), 184–193. <https://doi.org/10.1037/a0017169>
- Fagot, J., & Thompson, R. K. R. (2011). Generalized relational matching by guinea baboons (*Papio papio*) in two-by-two-item analogy problems. *Psychological Science*, 22(10), 1304–1309. <https://doi.org/10.1177/0956797611422916>
- Fagot, J., & Tomonaga, M. (1999). Global and local processing in humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*): Use of a visual search task with compound stimuli. *Journal of Comparative Psychology*, 113(1), 3–12.
- Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41, 1–63. https://doi.org/10.1207/s15327868ms0702_6
- Galesic, M., Walkyria Goode, A., Wallsten, T. S., & Norman, K. L. (2018). Using Tversky’s contrast model to investigate how features of similarity affect judgments of likelihood. *Judgment and Decision Making*, 13(2), 163–169.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170. <https://doi.org/10.1054/arth.2001.25504>
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52(1), 45–56. <https://doi.org/10.1037/0003-066X.52.1.45>
- George, D. N., & Pearce, J. M. (2003). Discrimination of structure: II. Feature binding. *Journal of Experimental Psychology: Animal Behavior Processes*, 29(2), 107–117. <https://doi.org/10.1037/0097-7403.29.2.107>
- George, D. N., Ward-Robinson, J., & Pearce, J. M. (2001). Discrimination of structure: I. Implications for connectionist theories of discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 27(3), 206–218. <https://doi.org/10.1037/0097-7403.27.3.206>
- Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75(4), 1098–1114. <https://doi.org/10.1111/j.1467-8624.2004.00728.x>
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117(3), 227–247. <https://doi.org/10.1037/0096-3445.117.3.227>

- Goldstone, R. L. (1994a). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 3–28. <https://doi.org/10.1037/0278-7393.20.1.3>
- Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52(2), 125–157. [https://doi.org/10.1016/0010-0277\(94\)90065-5](https://doi.org/10.1016/0010-0277(94)90065-5)
- Goldstone, R. L., & Medin, D. L. (1994). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 29–50. <https://doi.org/10.1037/0278-7393.20.1.29>
- Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin*, 132(6), 920–945. <https://doi.org/10.1037/0033-2909.132.6.920>
- Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, 25(6), 475–492. <https://doi.org/10.1016/j.tics.2021.01.006>
- Hahn, U. (2014). Similarity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(3), 271–280. <https://doi.org/10.1002/wcs.1282>
- Hahn, U., & Bailey, T. M. (2005). What makes words sound similar? *Cognition*, 97(3), 227–267. <https://doi.org/10.1016/j.cognition.2004.09.006>
- Hahn, U., & Chater, N. (1997). Concepts and similarity. In *Knowledge, concepts, and categories (studies in cognition)* (pp. 43–92).
- Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, 87, 1–32. <https://doi.org/10.1016/S01011111/j.1467-9280.2009.02310.x>
- Hall, G., & Rodríguez, G. (2017). Habituation and conditioning: Salience change in associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, 43(1), 48–61. <https://doi.org/10.1037/xan0000129>
- Hamel, L. (2009). Knowledge discovery with support vector machines. In *Knowledge discovery with support vector machines*. Wiley. <https://doi.org/10.1002/9780470503065>.
- Hau, D. B. M., & Call, J. (2009). Great apes' capacities to recognize relational similarity. *Cognition*, 110(2), 147–159. <https://doi.org/10.1016/j.cognition.2008.10.012>
- Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, 21(1), 101–148. <https://doi.org/10.1017/S0140525X98000703>
- Heyes, C. M. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, 5(6), 253–261.
- Hodgetts, C. J., & Hahn, U. (2012). Similarity-based asymmetries in perceptual matching. *Acta Psychologica*, 139(2), 291–299. <https://doi.org/10.1016/j.actpsy.2011.12.003>
- Hodgetts, C. J., Hahn, U., & Chater, N. (2009). Transformation and alignment in similarity. *Cognition*, 113(1), 62–79. <https://doi.org/10.1016/j.cognition.2009.07.010>
- Holmes, N. M., Chan, Y. Y., & Westbrook, R. F. (2019). A combination of common and individual error terms is not needed to explain associative changes when cues with different training histories are conditioned in compound: A review of Rescorla's compound test procedure. *Journal of Experimental Psychology: Animal Learning and Cognition*, 45(2), 242–256. <https://doi.org/10.1037/xan0000204>
- Hopkins, W. D., & Washburn, D. A. (2002). Matching visual stimuli on the basis of global and local features by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *Animal Cognition*, 5(1), 27–31. <https://doi.org/10.1007/s10071-001-0121-8>
- Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the massive memory picture database. *PLoS One*, 9(11). <https://doi.org/10.1371/journal.pone.0112644>
- Huber, L., & Lenz, R. (1993). A test of the linear feature model of polymorphous concept discrimination with pigeons. *The Quarterly Journal of Experimental Psychology: Section B*, 46(1), 1–18.
- King, M. L., Groen, I. I. A., Steel, A., Kravitz, D. J., & Baker, C. I. (2019). Similarity judgments and cortical visual responses reflect different properties of object and scene categories in naturalistic images. *NeuroImage*, 197(March), 368–382. <https://doi.org/10.1016/j.neuroimage.2019.04.079>
- Kirkpatrick-Steger, K., Wasserman, E. A., & Biederman, I. (1998). Effects of geon deletion, scrambling, and movement on picture recognition in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 24(1), 34–46.
- Larkey, L. B., & Markman, A. B. (2005). Processes of similarity judgment. *Cognitive Science*, 29(6), 1061–1076. https://doi.org/10.1207/s15516709cog0000_30
- Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta 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)
- Lawson, R., & Jolicoeur, P. (2003). Recognition thresholds for plane-rotated pictures of familiar objects. *Acta 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)
- Leech, R., Mareschal, D., & Cooper, R. P. (2007). Relations as transformations: Implications for analogical reasoning. *The Quarterly Journal of Experimental Psychology*, 60(7), 897–908. <https://doi.org/10.1080/17470210701288599>
- Luzardo, A., Alonso, E., & Mondragón, E. (2017). A Rescorla-Wagner drift-diffusion model of conditioning and timing. *PLoS Computational Biology*, 13(11). <https://doi.org/10.1371/journal.pcbi.1005796>
- Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, 32(4), 517–535. <https://doi.org/10.1006/jmla.1993.1027>
- Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and Cognition*, 24(2), 235–249. <https://doi.org/10.3758/BF03200884>
- Martin-Ordas, G., Berntsen, D., & Call, J. (2013). Memory for distant past events in chimpanzees and orangutans. *Current Biology*, 23(15), 1438–1441. <https://doi.org/10.1016/j.cub.2013.06.017>
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100(2), 254–278. <https://doi.org/10.1037/0033-295X.100.2.254>
- Mendes, N., Rakoczy, H., & Call, J. (2011). Primates do not spontaneously use shape properties for object individuation: A competence or a performance problem? *Animal Cognition*, 14(3), 407–414. <https://doi.org/10.1007/s10071-010-0375-0>
- Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, 4(MAR), 128. <https://doi.org/10.3389/fpsyg.2013.00128>
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 104–114. <https://doi.org/10.1037/0278-7393.10.1.104>
- Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97(2), 185–200. <https://doi.org/10.1037/0033-295X.97.2.185>
- Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, 101(4), 587–607. <https://doi.org/10.1037/0033-295X.101.4.587>
- Pearce, J. M. (2008). *Animal learning and cognition: An introduction*. Taylor & Francis.
- Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31(2), 109–178. <https://doi.org/10.1017/S0140525X08003543>
- Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology*, 58, 97–118. <https://doi.org/10.1146/annurev.psych.58.110405.085555>
- Poldrack, R. A. (2020). The physics of representation. *Synthese*, 199(1), 1307–1325. <https://doi.org/10.1007/s11229-020-02793-y>
- Rasmussen, A., Zucca, R., Johansson, F., Jirnhed, D. A., & Hesselow, G. (2015). Purkinje cell activity during classical conditioning with different conditional stimuli explains central tenet of Rescorla-Wagner model. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45), 14060–14065. <https://doi.org/10.1073/pnas.1516986112>
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In H. A. Black, & W. F. Prokasy (Eds.), *21. Classical conditioning II current research and theory* (pp. 64–99). Appleton-Century-Crofts. <https://doi.org/10.1101/gr.110528.110>
- Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16(3), 371–416. [https://doi.org/10.1016/0010-0285\(84\)90014-8](https://doi.org/10.1016/0010-0285(84)90014-8)
- Santos, L. R., Miller, C. T., & Hauser, M. D. (2003). Representing tools: How two non-human primate species distinguish between the functionally relevant and irrelevant features of a tool. *Animal Cognition*, 6(4), 269–281. <https://doi.org/10.1007/s10071-003-0171-1>
- Santos, L. R., Sulkowski, G. M., Spaepen, G. M., & Hauser, M. D. (2002). Object individuation using property/kind information in rhesus macaques (*Macaca mulatta*). *Cognition*, 83(3), 241–264. [https://doi.org/10.1016/S0010-0277\(02\)00006-9](https://doi.org/10.1016/S0010-0277(02)00006-9)
- Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22(4), 325–345. <https://doi.org/10.1007/BF02288967>
- Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210(4468), 390–398. <https://doi.org/10.1126/science.210.4468.390>
- Shepherd, J. (2018). Intelligent action guidance and the use of mixed representational formats. *Synthese*, 198(s17), 4143–4162. <https://doi.org/10.1007/s11229-018-1892-7>
- Simons, J. S., Lee, A. C. H., Graham, K. S., Verfaellie, M., Koutstaal, W., Hodges, J. R., ... Budson, A. E. (2005). Failing to get the gist: Reduced false recognition of semantic associates in semantic dementia. *Neuropsychology*, 19(3), 353–361. <https://doi.org/10.1037/0894-4105.19.3.353>
- Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General*, 133(3), 398–414. <https://doi.org/10.1037/0096-3445.133.3.398>
- Taylor, E. G., & Hummel, J. E. (2009). Finding similarity in a model of relational reasoning. *Cognitive Systems Research*, 10(3), 229–239. <https://doi.org/10.1016/j.cogsys.2008.09.004>
- Theves, S., Fernandez, G., & Doeller, C. F. (2019). The hippocampus encodes distances in multidimensional feature space. *Current Biology*, 29(7), 1226–1231. e3 <https://doi.org/10.1016/j.cub.2019.02.035>
- Tomasello, M., & Call, J. (2007). *Primate cognition*. Oxford University Press.
- Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states—The question is which ones and to what extent. *Trends in Cognitive Sciences*, 7(4), 153–156. [https://doi.org/10.1016/S1364-6613\(03\)00035-4](https://doi.org/10.1016/S1364-6613(03)00035-4)
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Origin of cultural cognition. *Behavioral and Brain Sciences*, 28, 675–735.
- Toussaint, G. T., Matthews, L., Campbell, M., & Brown, N. (2012). Measuring musical rhythm similarity: Transformation versus feature-based methods. *Journal of Interdisciplinary Music Studies*, 6(1), 23–53. <https://doi.org/10.4407/jims.2012.12.002>

- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(2), 127–190. <https://doi.org/10.1037/0033-295X.84.4.327>
- Vonk, J. (2003). Gorilla (*Gorilla gorilla gorilla*) and orangutan (*Pongo abelii*) understanding of first- and second-order relations. *Animal Cognition*, *6*(2), 77–86. <https://doi.org/10.1007/s10071-003-0159-x>
- Wagner, A. R., & Brandon, S. E. (2000). A componential theory of Pavlovian conditioning. In R. R. Mowrer, & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 23–64). Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410600691>
- White, J. B. (2008). Self-other similarity judgment asymmetries reverse for people to whom you want to be similar. *Journal of Experimental Social Psychology*, *44*(1), 127–131. <https://doi.org/10.1016/j.jesp.2006.10.015>
- Zentall, T. R., Wasserman, E. A., & Urcuioli, P. J. (2014). Associative concept learning in animals. *Journal of the Experimental Analysis of Behavior*, *101*(1), 130–151. <https://doi.org/10.1002/jeab.55>