

WHY ARE OUR SIMILARITIES SO DIFFERENT? A REPLY TO HUMPHREYS AND RIDDOCH

Keith R Laws¹ and Tim M Gale^{2,3}

(¹Brain and Cognition Research Group, Division of Psychology, Nottingham Trent University, UK; ²Departments of Computer Science and Psychology, University of Hertfordshire, UK; ³Queen Elizabeth Hospital, Welwyn Garden City, UK)

Humphreys and Riddoch (2002: hereafter, H&R) entitle their commentary on Laws and Gale (2002) with the following question: “Do pixel level analyses describe psychological perceptual similarity”? (We will use the term perceptual similarity to refer also to structural similarity as used by Humphreys and colleagues). This perhaps betrays some misunderstanding of our intention. Throughout our paper, we specifically refer to *visual* overlap – which we take to mean retinotopic similarity at the pixel level. This differs from notions of psychological or perceptual similarity and certainly differs from the position advanced by Humphreys and colleagues. Although we do not view the two approaches as mutually exclusive, we have some reservations about the utility of perceptual similarity as measured by Contour Overlap (CO) and partonomic features (Humphreys et al., 1988). Moreover, on grounds of parsimony, it is important to examine the role of *low-level* variables in object recognition and category specificity before turning to *high-level* variables i.e. psychological/perceptual variables. In response to the commentary by H&R (2002), we would like to draw attention to some issues that relate to: (a) the points they raise about Euclidean Overlap (EO); and (b) difficulties with their conception of perceptual similarity.

DIRECT COMPARISONS OF EO AND CO

As H&R point out, EO accounted for 4.4% of the variance in the naming errors made by 32 normal subjects to the whole Snodgrass and Vanderwart corpus. Although we did not report correlation values for CO in our paper, the variance in naming errors predicted by CO is much lower at only 0.1%. Additionally, we (Laws et al., 2002a) have found that EO significantly predicts picture-naming latencies in normal subjects (also \approx 4% of the variance: this level of correlation is similar or better than that reported for other ‘visual’ variables such as visual complexity). In multiple regression analyses, EO was a significant predictor (along with age of acquisition) while CO was not (neither were other typical variables such as familiarity, visual complexity and name frequency). Moreover, in the same study, EO significantly predicted naming of the same pictures when these were covered with masks of different densities (a random-noise filter covering 10% and 20% of each picture with black pixels); however, again CO was not a significant predictor. Hence, when directly pitted

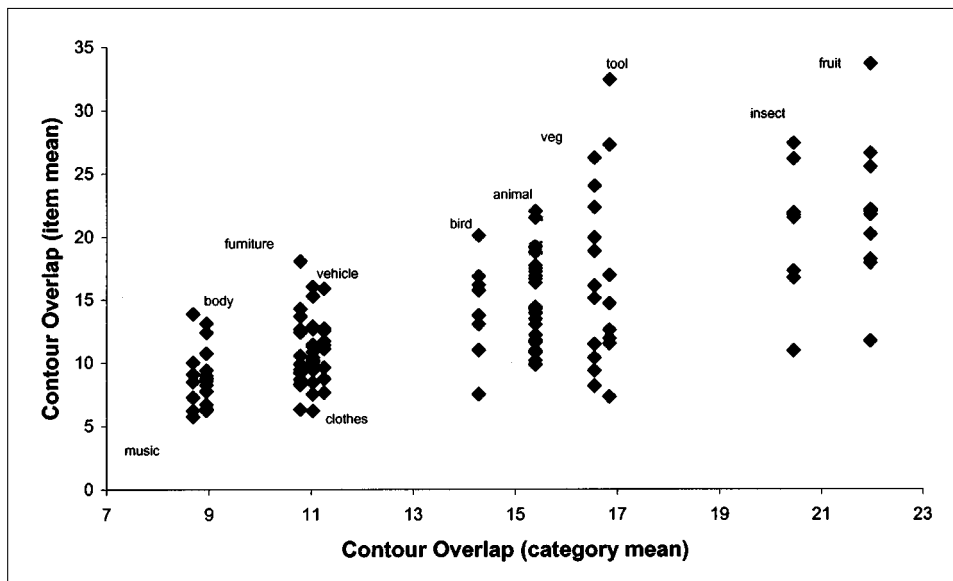


Fig. 1 – The distribution of contour overlap scores (from Humphreys et al., 1988).

against each other in these two studies, EO outperforms CO in predicting both normal errors and normal latencies.

Another reason why we consider EO to be an important metric is that it is the *only* extant variable (visual, linguistic or otherwise) that predicts the association of body parts with nonliving things, and musical instruments with living things (Laws et al., 2002b), which has been documented in category specific patients. By contrast, if we plot the Contour Overlap data (from Humphreys et al., 1988), this places animals and musical instruments at different ends of the visual overlap continuum; CO clusters musical instruments and body parts together as having the lowest overlap of all categories; and tools as having higher overlap than animals, birds and vegetables (see Figure 1: cf. Figure 3 from Laws and Gale, 2002; see also Laws et al., 2002b). Therefore, although H&R question whether analyses at a pixel level are appropriate, the fact is that EO arguably predicts this particular behavioural pattern better than other measures.

MEASURING EO AND CO: HOW DO THEY DIFFER?

The suggestion that, relative to EO, CO is “abstracted to some degree from the original images” (p. 2) is difficult to accept since we cannot see how any qualitative distinction can be made between pixel-based metrics (such as EO) and *any* length-based metric (such as CO). Both measures normalise for size and so the only differences are that CO: (i) excludes internal detail and (ii) normalises for orientation. Even with internal details missing, CO is still derived

from the same *single* set of pictures. While removing internal line detail might arguably reduce visual complexity, the way that object outlines are drawn is as susceptible to artistic whim as the way that internal features are drawn. Moreover, normalising for orientation can be problematic. It is easy to normalise some categories (e.g. animals) where there is an easily discernible ‘front end’ or ‘head’, but more difficult to normalise non-living things (e.g. furniture, tools, and musical instruments) and even some living things (fruit and vegetables). In the latter examples, any chosen orientation is essentially arbitrary and so this could arguably influence visual overlap (whether measured by EO or CO). Furthermore, the vast majority of studies that have used this particular set of stimuli have presented them in their standard orientations and so it would seem useful to present data on those particular orientations.

Unlike previous studies, however, we have explicitly investigated orientation as a factor. To do this, we examined EO for silhouette (EO[sil]) and greyscale (EO[greyscale]) versions of the Snodgrass and Vanderwart images that were standardised for size. By comparison with the EO measure (EO[line]) reported by Laws and Gale (2002), internal detail was removed completely under EO[sil], and increased under EO[greyscale] (see Figure 2). These new EO measures were derived for the same 141 items, using the original *and* reoriented versions. Some items were realigned using the same left-right orientation e.g. all animal heads were reoriented to the same direction. Where subcategories had no obvious canonical orientation, we made arbitrary but consistent decisions (e.g. tools had handles pointing top left).

For EO[sil], we found no overall living/non-living difference (0.60 vs. 0.605, $F = 1.15$, $p = 0.28$). Reorientation increased overlap ($F_{1, 139} = 6.9$, $p = .01$), and the interaction with category approached significance ($F_{1, 139} = 3.14$, $p = .07$) – increasing overlap for *nonliving* rather than living things. Turning to the more detailed set of images in greyscale, EO[greyscale] revealed no overall living/non-living difference ($F < 1$: 0.57 vs. 0.56). Reorientation significantly increased

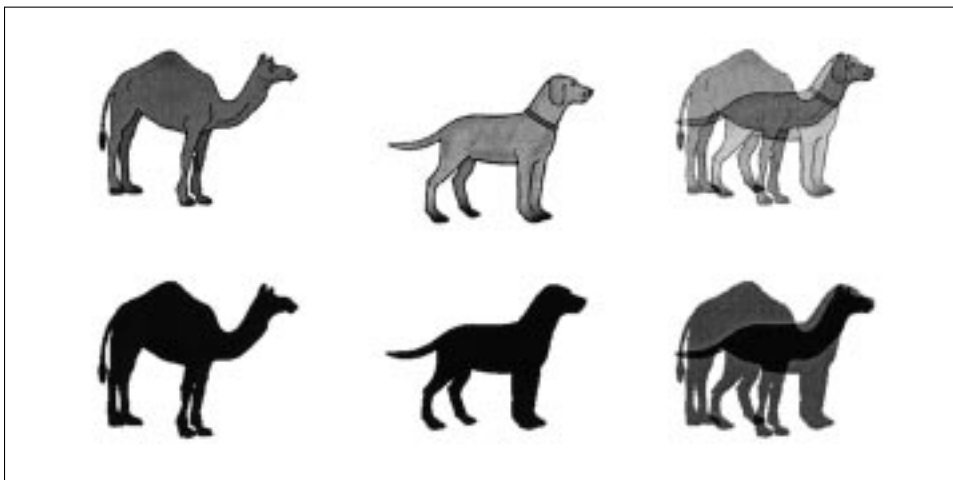


Fig. 2 – Examples of retinotopic overlap as measured by EO[greyscale] and EO[sil].

overlap ($F_{1,139} = 6.9$, $p = .01$) and again the interaction between orientation and category approached significance ($F_{1, 139} = 3.13$, $p = .08$), with an increase in overlap for nonliving things once more. Removal of body parts and musical instruments from each analysis did not alter the findings. The main conclusions from this work are that: (a) neither EO for silhouettes nor greyscale versions of the same stimuli distinguish living from nonliving things; (b) category does not interact with orientation and if anything, similarity may increase more for nonliving things after reorientation; and finally (c) visual overlap for simple outline plus internal detail, as described by Laws and Gale (2002) seems to best distinguish between living and nonliving things (see Laws et al., in press for a direct comparison).

One final way in which CO and EO differ relates to the composition of superordinates. It is notable that the composition of superordinates used by Humphreys et al. (1988) when devising CO seems overly inclusive especially for some nonliving categories. For example: <Toys> included: pram, bell, book, box, boat, telephone, whistle; a category referred to simply as <Objects> included: pipe, jug, cigarette, coat-hanger, cotton-reel, saltshaker, candle, lock, key, pen; and <Household> included: bag, watering can, comb, plug, glasses, clothes-peg, cigar, toothbrush. These are, essentially, *ad-hoc* groupings that may well lower visual-categorical coherence for nonliving things. By contrast, EO was derived from a subset of 11 semantic categories that have been widely used and reported in the neuropsychological literature (see Laws and Gale, 2002).

Finally, H&R raise concerns that EO is only predictive when it is calculated with respect to superordinate categories and that the predictive value disappears when calculated across all items. The alternative is to compare any object with every other possible object (or at least a large set of all possible objects), where it is, of course, very unlikely that there will be any overall visual coherence. The central point, however, about visual overlap is that it does relate to semantic categories: indeed, the very issue is the extent to which certain semantic categories are rendered more 'coherent' by internal visual similarity. Moreover, it is difficult to see how H&R's CO measure differs from EO in this respect: it too is based only on within-category comparisons and it would be meaningless to measure it in any other way.

DOES THE H&R NOTION OF STRUCTURAL SIMILARITY PREDICT BEHAVIOURAL DATA?

Although H&R claim that, "...behavioural measures do not constrain the kinds of perceptual information being used", they point to studies where they suggest that "measures such as overlap, in stated parts and the percentage contour overlap also generate results that match findings when behavioural measures of similarity are derived" (p. 2). Indeed, Humphreys and colleagues have claimed an essentially isomorphic mapping between perceptual similarity-dissimilarity and living-nonliving. For example, Humphreys et al. (1988) include tools amongst structurally distinct items even though they have greater CO than animals, birds and vegetables; but on the other hand, exclude musical

instruments from the structurally dissimilar even though they show the least CO: see Figure 1). Moreover, studies reporting worse behavioural performance with living than nonliving things do not necessarily represent evidence for structural similarity being the important underlying factor. There are many reasons for this, including: (a) the mapping is not isomorphic (see Figure 1); (b) the logic of this inference is not valid i.e. any behavioural differences on living and nonliving things may reflect a host of variables other than structural similarity; (c) alternative explanations have not been eliminated in some studies e.g. visual complexity and familiarity differences (see Laws, 2001); and (d) a body of evidence demonstrates that normal subjects are quicker and more accurate at naming living things (see below).

To take one of H&R's examples, the widely cited study of Gaffan and Heywood (1993) who reported that monkeys made more errors when distinguishing between pairs of living than pairs of nonliving items (from the Snodgrass and Vanderwart corpus). H&R's interpretation of the monkey's performance is that it "*prima facie* seem(s) to be determined by the perceptual similarity between items...living things fall on the side of being more similar than non-living things". This inference simply affirms the consequence and is not valid (whatever the evidence). While Humphreys and colleagues argue that greater perceptual similarity for living than nonliving things may lead to the prediction of greater difficulty recognising living things (though see below); the reverse hypothesis does not follow i.e. that greater difficulty with living things implies perceptual similarity is responsible. It is also difficult to see how these data could provide unequivocal support for H&R's notion of perceptual similarity. If, as H&R argue, perceptual similarity is (or should be) *abstracted*, then what exactly is underlying the performance of the monkeys? Can they really be abstracting common 'parts' or partonomic information? A simpler explanation is that Gaffan and Heywood did not match across the living and nonliving category for visual complexity and so, the living things were simply more difficult to distinguish because they are more visually complex.

BEHAVIOURAL DATA THAT CONTRADICTS H&R'S NOTION OF PERCEPTUAL SIMILARITY

As noted above, a direct prediction of the perceptual similarity argument proposed by H&R is that living things should be more difficult to recognise and name for normal subjects. Indeed, studies by Humphreys and colleagues have documented better naming of structurally distinct (exclusively living things) than structurally similar (exclusively nonliving) things (Humphreys et al., 1988; Vitkovitch et al., 1993; Lloyd-Jones and Humphreys, 1997). By contrast, we have consistently found better naming of living than nonliving things by normal subjects (Laws, 2002; Laws, 2000; Laws, 1999; Laws et al., 2002a) when stimuli are matched across category for variables known to affect naming (i.e. familiarity, name frequency, visual complexity).

In this context, an attempt to reconcile these two approaches has suggested that task demands may influence whether living or nonliving advantages occur

in normal subjects (see Gerlach, 2001). Specifically, Gerlach (2001: see also Humphreys and Forde, 2001) have proposed that the use of *degraded* viewing conditions (e.g. 20 msec presentation time: Laws and Neve, 1999) puts greater reliance on global visual processing and that this benefits living things; and that by contrast, the use of undegraded stimuli (e.g. Humphreys et al., 1988) permits local information to be more fully processed, and this benefits the processing of nonliving things. The argument is that stimulus degradation results in a loss of high spatial frequency information, that makes perceptual structures less reliable (because the living things show less variation in base-level exemplars: see Laws and Neve, 1999) and therefore, global shape is processed at the expense of local detail. Given their assumption that living things share greater global shape, these authors (Gerlach, 2001; Humphreys and Forde, 2001) argue that global shape may reveal more of an items identity when the item is a natural object than a man-made artefact. Nevertheless, the living thing advantage that we have reported has been documented in several paradigms using both degraded and undegraded stimuli, for example: naming-to-deadline (Laws, 2000); naming latency (Laws, 1999); naming pictures of graded difficulty, presented undegraded and with no time or other constraints (Laws, 2002); and critically, in the only study that has directly compared naming latencies from the same subjects to the same images when undegraded and degraded at different levels (Laws et al., 2002a)¹.

H&R'S CRITICISMS OF EO (AS A PIXEL-BASED MEASURE)

H&R point out that other studies (Tranel et al., 1997; Gale et al., 2001) taking measures at a pixel level fail to match those in Laws and Gale (2002). To deal briefly with the Tranel study, their reported pattern did not distinguish living from nonliving things, but more importantly also fails to accord with the position of H&R (see Figure 1). Turning to the neural network study of Gale et al (2001), although this work also fails to accord with H&R's position e.g. see the profiles for musical instruments and animals for CO (Figure 1), we will deal with this in more detail. In Gale et al's (2001) simulations, animals and musical instruments generated more tightly clustered Self-Organising Feature Map (SOFM) representations than furniture and clothing, indicating greater 'perceptual' crowding for the former categories. We have now extended this work and replicated this finding using a new set of greyscale variants of the Snodgrass and Vanderwart (1980) corpus (see examples in Figure 2). As before, animals *and* musical instruments were more tightly clustered than furniture and clothing ($F_{1, 36} = 4.70, p < 0.05$); however, when we examined the same 11 categories that were used by Laws and Gale, no overall living/nonliving difference emerged ($F < 1$). These data clearly show that visual similarity can give rise to category effects, but that this pattern does not, overall, honour the living/nonliving distinction. In these SOFM simulations, alternative explanations

¹ It remains possible that such effects apply to object decision tasks (see Gerlach, 2001); however, this hypothesis requires examination using the same stimuli (i.e. Gerlach used different sets for his object decision tasks)

concerning visual complexity or other psychological variables can be eliminated, and so it is possible to conclude that any emerging category effects derive from *visual* similarity (cf. studies with human or primate subjects). Moreover, an important difference between the SOFM measures and EO is that the former are reduced-dimensional *abstractions* of pixel representations and, as such, are more akin to internal or ‘perceptual’ representations than retinotopic maps². Indeed, a SOFM forms a 2-dimensional surface in hyper-space that most closely approximates the topology of the training set. Reducing the number of dimensions in which visual information is represented inevitably means that not all information is preserved and so it is difficult to specify the basis of encoded similarity exactly. Therefore, we would interpret the SOFM data of Gale et al. (2001) as being more akin to what H&R refer to as a ‘perceptual measure’. Whether or not this is a good model of human perceptual representation may be open to debate but, minimally, one can argue that this kind of representation is an abstraction, and not an analogue of a pixel representation.

Furthermore, in our replication work, the SOFM measures derived from *greyscale* variants of the Snodgrass and Vanderwart (1980) stimuli correlated significantly with the EO *line* measures (from Laws and Gale, 2002). This again suggests that EO plays an important role in object recognition (see Laws et al., in press). First, because EO correlates with a truly *abstracted* measure i.e. the SOFM measure of overlap. The SOFM representations are *abstracted* in the sense that they develop in response to bottom-up visual information and are not influenced by semantic relationships (cf. ratings from human subjects: Humphreys et al., 1995). They are *abstracted* also in the sense that the SOFM re-represents the information in a reduced number of dimensions. Second, EO was derived from the original line-drawn images, while the SOFM measure was derived from greyscale versions of the same images (i.e. and so, contained internal information such as shading and surface texture that were not available in the originals) and so, the SOFM appears to emphasise the outline *and* the internal detail. By contrast, CO did not correlate significantly with the SOFM measure. So, none of our work supports the view that living things are, *overall*, more visually crowded than nonliving things at either a retinotopic or an abstracted ‘perceptual’ level; using line drawings, greyscale or silhouettes.

In conclusion, we suggest that retinotopic variables – such as EO – do have predictive validity for object naming tasks, both under normality and under pathology. Such measures may not generalise beyond the Snodgrass and Vanderwart (1980) set of pictures, but we have never claimed otherwise (and the same is of course true for other well-accepted predictor variables such as visual complexity). Nonetheless, given that the vast majority of category specific studies (we estimate 90%+) have been based upon the Snodgrass and Vanderwart stimuli, we do not view this as a limitation to the utility of EO.

Acknowledgements. We would like to thank various people who have contributed to this research, including Verity Leeson, Neil Davey, Ray Frank, and Christine Neve. Thanks

² This issue is important since it is unclear whether Humphreys and colleagues are referring to similarity of input stimuli or for stored representations or both

to Professor Humphreys and colleagues for generously allowing us access to their Contour Overlap data and to Professor Della Sala for his helpful comments on an earlier draft of this paper.

REFERENCES

- GAFFAN D and HEYWOOD CA. A spurious category-specific visual agnosia for living things in normal human and non-human primates *Journal of Cognitive Neuroscience*, 5: 118-128, 1993.
- GALE TM, DONE DJ and FRANK RJ. Visual crowding and category-specific deficits for pictorial stimuli: a neural network model *Cognitive Neuropsychology*, 18: 509-550, 2001.
- GERLACH C. Structural similarity causes different category-effects depending on task characteristics. *Neuropsychologia*, 39: 895-900, 2001.
- HUMPHREYS GW and FORDE EME. Hierarchies, similarity and interactivity in object recognition: On the multiplicity of 'category-specific' deficits in neuropsychological populations. *Behavioural and Brain Sciences*, 24: 453-509, 2001.
- HUMPHREYS GW, LAMOTE C and LLOYD-JONES TJ. An interactive activation approach to object processing: Effects of structural similarity, name frequency, and task in normality and pathology. *Memory*, 3: 535-586, 1995.
- HUMPHREYS GW and RIDDOCH J. Do pixel-level analyses describe psychological perceptual similarity? A comment on 'Category-specific naming and the 'visual' characteristics of line drawn stimuli' by Laws and Gale. *Cortex*, 38: 3-5, 2002.
- HUMPHREYS GW, RIDDOCH J and QUINLAN PT. Cascade processes in picture identification. *Cognitive Neuropsychology*, 5: 67-103, 1988.
- KOHNEN T. *Self-Organisation and Associative Memory*. Berlin: Springer-Verlag, 1988.
- LAWS KR. Gender affects latencies for naming living and nonliving things. *Cortex*, 35: 729-733, 1999.
- LAWS KR. Category-specific naming errors in normal subjects: The influence of evolution and experience. *Brain and Language*, 75: 123-133, 2000.
- LAWS KR. What is structural similarity and is it greater in living things? *Behavioural and Brain Sciences*, 24: 486-487, 2001.
- LAWS KR. Category-specific naming and modality-specific imagery. *Brain and Cognition*, 48: 418-420, 2002.
- LAWS KR and GALE TM. Category-specific naming and the 'visual' characteristics of line drawn stimuli. *Cortex*, 38: 7-21, 2002.
- LAWS KR, GALE TM, FRANK RJ and DAVEY N. Visual Similarity is greater for line drawings of nonliving than living things: The importance of musical instruments and body-parts. *Brain and Cognition*, 48: 421-424, 2002b.
- LAWS KR, GALE TM, LEESON VC and FRANK RJ. The influence of surface and edge-based visual similarity on object recognition. *Brain and Cognition*, in press.
- LAWS KR, LEESON VC and GALE TM. The effect of 'masking' on picture naming latencies. *Cortex*, 38: 137-147, 2002a.
- LAWS KR and NEVE C. A 'normal' category-specific advantage for naming living things. *Neuropsychologia*, 37: 1263-1269, 1999.
- LLOYD-JONES T and HUMPHREYS GW. Perceptual differentiation as a source of contour effects in object processing: evidence from naming and object decision. *Memory and Cognition*, 25: 18-35, 1997.
- SNODGRASS JG and VANDERWART M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6: 174-215, 1980.
- TRANIEL D, LOGAN CG, FRANK RJ and DAMASIO AR. Explaining category-related effects in the retrieval of conceptual and lexical knowledge for concrete entities: operationalization and analysis of factors. *Neuropsychologia*, 35: 1329-1339, 1997.
- VITKOVITCH M, HUMPHREYS GW and LLOYD-JONES T. On naming a giraffe a zebra: Picture naming errors across different object categories. *Journal of Experimental Psychology: Learning Memory and Cognition*, 19: 243-259, 1993.