Jake Quilty-Dunn

Iconicity and the Format of Perception[1]

Abstract: *According to one important proposal, the difference between perception and cognition consists in the representational formats used in the two systems (Carey, 2009; Burge, 2010; Block, 2014). In particular, it is claimed that perceptual representations are iconic, or image-like, while cognitive representations are discursive, or language-like. Taking object perception as a test case, this paper argues on empirical grounds that it requires discursive label-like representations. These representations segment the perceptual field, continuously pick out objects despite changes in their features, and abstractly represent high-level features, none of which appears possible for purely iconic representations.*

1. Introduction

The difference between perception and thought is intuitive but poorly understood. Are perception and cognition distinct mental systems? If so, what distinguishes them? According to one important proposal, the difference at least partly consists in the representational formats used in the two systems (Carey, 2009; Burge, 2010; Block, 2014). In particular, representations in perception and perceptual imagery are

Correspondence:
Jake Quilty-Dunn, Cognitive Science and Philosophy, The Graduate Center, CUNY, New York, USA. *Email: jamquid@gmail.com*

[1] Editorial note: This paper was the winner of the Annual Essay Prize at the Center for Philosophical Psychology, University of Antwerp, in 2016. The *Journal of Consciousness Studies* is collaborating with the Prize to publish the winners of this competition in our journal. Please see http://webh01.ua.ac.be/bence.nanay/paw.html for further details.

taken to be iconic, or image-like, while representations in cognition are taken to be discursive, or language-like.

Iconic representations (or 'icons') obey the following principle:

> (ICONICITY): Every part of the representation represents some part of the scene represented by the whole representation.

Though it may not be true of every kind of image-like representation, ICONICITY intuitively applies to paradigm cases like photographs (Quilty-Dunn, unpublished), and it describes the notion of iconic mental representation that typically figures in cognitive science (Kosslyn, 1980, p. 33; 1994, p. 5; Johnson-Laird, 2006, p. 25; Fodor, 2007, p. 108; Carey, 2009, p. 452). Icons have been posited to explicate so-called 'iconic memory' systems (Sperling, 1960; Bradley and Pearson, 2012), the depictive representations posited by the quasi-pictorialist side of the imagery debate (Kosslyn, 1980; 1994; Pearson *et al.*, 2015), and mental models in the psychology of reasoning (Johnson-Laird, 2006). Visual imagery is plausibly iconic (*pace* Pylyshyn, 2003). Given the extensive neuropsychological commonalities between imagery and early perceptual systems (Pearson *et al.*, 2015), therefore, there is good reason to suppose that at least some representations in vision have an iconic format.

Since every part of an icon represents some part of the represented scene, there is no way of carving up the representation such that the carved-up parts are not representational. This syntactic freedom of icons contrasts with discursive representations, such as sentences, which have parts that are not meaningful at all. For example, arbitrarily combining the first and last word of a sentence or carving a simple word in half will not typically yield meaningful representational parts that contribute to the meaning of the sentence. Discursive representations thus have canonical decompositions into privileged parts, hereafter 'constituents', or else are simple, compositionally efficacious constituents that lack interpretable parts altogether, such as individual words or (arguably) atomic concepts (Fodor, 2007). Without the aid of discursive representations to segment and label particular parts, there are no such privileged parts in an icon — iconic models of mental imagery, for example, posit associated discursive representations to label objects and describe parts of images (Kosslyn, 1980, pp. 145–7).

Consider the following claim:

(FORMAT): The difference between perception and cognition consists largely in the fact that representations in perception are iconic and representations in cognition are discursive.

Carey (2009), Burge (2010), and Block (2014) appear to endorse FORMAT, though they may wish to supplement it with some architectural border between perception and cognition. I will take FORMAT as a pure example of the general approach and examine its prospects. The test case I will use is object perception, which Burge and Carey agree is part of perception proper (Burge, 2010, pp. 437–70; Carey, 2011, p. 155) and thus should, according to FORMAT, be iconic. I will argue below that the data suggest otherwise.

2. Object Representations

While object recognition in high-level vision involves categorization and hence arguably involves the deployment of concepts, a more basic form of visual object perception involves simply tracking objects as such. Kahneman, Treisman and Gibbs (1992) showed the existence of an interesting effect of how we visually keep track of objects, which they called the object-specific preview benefit ('OSPB') — see Figure 1. Subjects were presented with two objects on a screen (such as squares or triangles) labelled, for example, 'P' and 'S'. The labels then disappeared and the objects moved in different directions. Then a letter would flash in one of the objects and subjects had to answer whether that letter was one of the original letters (e.g. 'P' and 'S') or another letter. Subjects were told it did not matter for the task which object the letter was originally flashed in. If, however, the letter was the same letter that flashed in that particular item originally, subjects were quicker at identifying it; e.g. if the 'P' appeared in the original 'P' object subjects were faster to recognize it than if that 'P' flashed in the original 'S' object. This was true even when controlling for location, shape, and other observable properties, so the observed 'preview benefit' is object specific. Kahneman *et al*. posited the existence of 'object files' to explain the OSPB. Object files are representations of particular objects that do not represent objects via any particular feature, but rather cluster features by attributing them to the same object.

Object files are linked to the 'visual indexes' or 'FINSTs' studied by Pylyshyn, Scholl, and others (e.g. Pylyshyn and Storm, 1988; Scholl, Pylyshyn and Franconeri, 1999; Pylyshyn, 2003; Haladjian and Pylyshyn, 2008). While the notion of object files arose from

studying the sometimes surprising ways features are clustered together in mid-level vision, the notion of visual indexes came from studying the equally surprising ways objects are tracked across space and time. In the basic multiple-object tracking ('MOT') paradigm, participants foveate on a fixation cross while some number of objects (e.g. eight squares) populate the rest of the stimulus. Some small number (e.g. four) of the objects will flash, indicating that they are to be tracked, while the rest are to be ignored. Participants are remarkably good at tracking up to four or five objects, even when the objects are qualitatively identical and intersect each other's paths and even when they change features or are occluded by hidden barriers (see Pylyshyn, 2003, for an overview). Participants cannot even report the spatio-temporal criteria used to individuate and track objects — though if motion ceases and an object disappears, subjects can say where it was and what direction it was headed in (Scholl, Pylyshyn and Franconeri, 1999), suggesting that 'participants have conscious access to a spatio-temporal address of a currently attended object, though not always to other features of the indexed object' (Carey, 2009, p. 75).

The relation between object files and visual indexes is not obvious. But they are plausibly part of one and the same capacity to segment the visual field into objects (Scholl and Leslie, 1999). Visual indexing is a mechanism of deploying attention to particular objects by location. Object files are the explicit representations that are the output of that mechanism, representations that refer to the particular objects and bind associated features by attributing them to the referents. When an object is perceived, then, visual attention is deployed and facilitates the construction of a representation that picks out the object, represents it as occupying a certain spatio-temporal position, and attributes various other features to it, including high-level ones (e.g. Jordan, Clark and Mitroff, 2010). Thus Kahneman *et al.* write that a visual index can be thought of as 'the initial spatiotemporal label that is entered in the object file and that is used to address it' and that a visual index 'might be the initial phase of a simple object file before any features have been attached to it' (Kahneman, Treisman and Gibbs, 1992, pp. 215–6; Scholl and Leslie, 1999). Aside from its theoretical elegance, this story is bolstered by the fact that, when subjects are given an MOT task where they track half of the visible objects, the OSPB is enhanced for the tracked objects (Haladjian and Pylyshyn, 2008).

This plausible account is *prima facie* incompatible with perception being wholly iconic. Object representations do precisely what icons

should not be able to do: they segment the visual field into discrete objects, each object representation standing for a particular individual. Objects 'pop out' of the visual field and become available for tracking. Icons do not come segmented, with certain depicted objects explicitly popping out of the represented scene; such popping out requires additional representational vehicles to segment the icon and represent the segmented objects. The fact that there is a distinct representational vehicle for each distinct object is why capacities for object perception are limited to a certain number of objects. For example, MOT fails above four or five objects (Pylyshyn, 2003, pp. 225–6). Doing the MOT tasks for yourself helps make this effect phenomenologically vivid.[2]

This phenomenon, which Fodor calls an 'item effect' (2007, p. 111), is precisely the sort of thing that should not happen if the relevant representations are iconic. An icon, such as a photograph, can represent two dozen objects as easily as three — indeed, Neisser (1967) originally coined the term 'iconic memory' to explain the famous Sperling (1960) results apparently showing that subjects perceptually encode more letters in a briefly presented array than they can conceptualize. It is because tracking more objects requires constructing more representations (indeed, exactly as many more) that difficulty increases with more objects and that MOT has a capacity limit determined by the number of objects rather than by their size or other features. Thus Henderson and Ames write that 'there is a cost associated with the construction' of each object file (1994, p. 836). Object perception appears to involve, therefore, a canonical decomposition of perceptual representations into constituents: distinct, discrete, discursive representations of individual objects.

3. Iconic Object Files?

To rebut this line of argument, one might follow Carey in suggesting that object files are individual icons (e.g. Carey, 2009, pp. 138, 149). As mentioned above, an iconic array is not segmented into representations of discrete objects. But perhaps each discrete object representation generated by segmentation processes is itself just another icon. For example, if I draw a picture of a particular red ball, then the

[2] Demonstrations are available at http://www.yale.edu/perception/Brian/demos/MOT.html.

drawing is an icon that apparently represents that particular object. Perhaps object files function similarly, without the aid of any discursive representations. That is, the perceptual field as a whole has a canonical decomposition into object representations, but each object representation is an icon that lacks a canonical decomposition and satisfies ICONICITY.

The range of properties an object file can represent renders this proposal mysterious. These properties include not only low-level properties such as shape and colour, but also which letter appears with the object even independently of typeface and case (Henderson, 1994), that the object is labelled as a fish even independently of whether the label is the word 'fish' or a picture of fish (Gordon and Irwin, 2000), and that the object is labelled as a piano even independently of whether the label is a picture of a piano or the sound of a piano being struck (Jordan, Clark and Mitroff, 2010). Moreover, given that the OSPB involves maintenance of information about features briefly presented moments before, the object representation must continually encode features that are not presently visible. It's hard to see how icons could explicitly represent abstract or absent features without aid from discursive interpretation. Since icons lack separate constituents for individual properties (e.g. the same syntactic part of an image of a red square represents both its colour and its shape), they cannot incorporate a discursive symbol that stands for a property and abstracts away from its low-level features. Indeed, the lack of discrete syntactic items standing for properties may limit icons to representing continuous magnitudes like colour and spatio-temporal properties (Kosslyn, 1980, p. 34; Quilty-Dunn, unpublished). Since an icon cannot incorporate a discursive symbol standing for (e.g.) *piano*, it does not seem possible for icons to explicitly encode the abstract features that object files do.

Furthermore, the analogy to a drawing of an object is problematic, since my drawing of a red ball is arguably of that particular ball because I had that particular ball in mind (i.e. had a discursive mental representation of it). If the perceptual object representation is simply an icon, then it lacks a distinct label-like representational element that picks out the object in addition to elements that represent it as having certain features (such as its size and colour). But object representations track objects even if they change all relevant features, including location. Thus Pylyshyn (2008) writes that visual indexes are akin to bare demonstratives, referring to their objects under no description at all (*cf.* Scholl and Leslie, 1999). Even if that analogy is mistaken, the

data call for a persistent label-like representation for each visually tracked object that selects its object and stores associated contents. This discursive label also functions to bind the associated contents together by enabling them to be attributed to its referent.

The proposal that object files are icons, therefore, faces a dilemma. If there are no discursive representations involved, then what is the object file over and above a certain cluster of iconic contents? If the answer is 'Nothing' then the proposal does not explain how object files work. In particular, it robs us of the explanation that representations of features X, Y, and Z are bound because an explicit object representation picks out an object and represents it as being X, Y, and Z. But that explanation seems like the only one on offer, so an account that gives it up is left without an explanation of how features are bound to objects in perception. If, on the other hand, there is a representational vehicle, over and above the iconic contents, serving to pick out the object and bind the feature representations, then it is not the *object file* that is iconic, but rather (at most) the information that is bound to it. On this horn, the proposal seems to be committed to a distinct, discrete representation of a particular distinct, discrete object — in other words, a discursive object representation.

Note that the issue at hand is not whether icons bind features *simpliciter*. Let it be the case that they do. The issue is whether they bind features of objects, to objects. That is, while it may be true that a picture binds colour, shape, and location properties (perhaps by attributing features to place-times), it will nevertheless fail to be true that it binds those properties under the description of their co-instantiation in a cohesive, bounded object. The present concern is that, in fact, icons lack the representational apparatus to bind features by picking out an object and attributing those features to the object. They might still bind those features (though the problem of abstract and absent features remains), but they cannot do so by attributing them to an object. If we suppose, as the empirical literature seems to force us to, that perceptual object representations have the representational apparatus to pick out objects and attribute features to them, then we have to suppose that perceptual object representations are not iconic. Thus perception is not (wholly) iconic.

One option for proponents of FORMAT is to reject ICONICITY as a characterization of icons and instead develop another characterization that is compatible with the data presented above. Since no defender of FORMAT has proposed such an alternative, however, this option remains unexplored. Its prospects also seem to be grim. For one thing,

redefining the key term to avoid objections seems *ad hoc*. Furthermore, the problem for FORMAT is not simply that perceptual object representations fail to be iconic in the standard cognitive-scientific sense. They function like labels that segment the perceptual field, continuously pick out objects despite changes in their features, and explicitly represent high-level features, like being a piano, in an abstract, amodal format. Such representations fail to be to picture-like in any intuitive sense. If there is a perception–cognition border, therefore, it is not due to the truth of FORMAT.

Acknowledgments

Thanks to Bence Nanay and the Centre for Philosophical Psychology, University of Antwerp. I am grateful to Ross Colebrook, Ryan DeChant, Nemira Gasiunas, E.J. Green, Grace Helton, Zoe Jenkin, Eric Mandelbaum, David Papineau, and especially Jesse Prinz for comments on earlier drafts.

References

Block, N. (2014) Seeing-as in the light of vision science, *Philosophy and Phenomenological Research*, **89** (3), pp. 560–572.

Bradley, C. & Pearson, J. (2012) The sensory components of high-capacity iconic memory and visual working memory, *Frontiers in Psychology*, **3** (355), doi: 10.3389/fpsyg.2012.00355.

Burge, T. (2010) *Origins of Objectivity*, Oxford: Oxford University Press.

Carey, S. (2009) *The Origin of Concepts*, Oxford: Oxford University Press.

Carey, S. (2011) Précis of *The Origin of Concepts*, *Behavioral and Brain Sciences*, **34**, pp. 113–167.

Fodor, J.A. (2007) The revenge of the given, in McLaughlin, B. & Cohen, J. (eds.) *Contemporary Debates in Philosophy of Mind*, pp. 105–116, Oxford: Blackwell.

Gordon, R.D. & Irwin, D.E. (2000) The role of physical and conceptual properties in preserving object continuity, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26** (1), pp. 136–150.

Haladjian, H. & Pylyshyn, Z. (2008) Object-specific preview benefit enhanced during explicit multiple object tracking, *Journal of Vision*, **8** (6), p. 497.

Henderson, J.M. (1994) Two representational systems in dynamic visual identification, *Journal of Experimental Psychology: General*, **123** (4), pp. 410–426.

Henderson, J.M. & Ames, M.D. (1994) Roles of object-file review and type priming in visual identification within and across eye fixations, *Journal of Experimental Psychology: Human Perception and Performance*, **20** (4), pp. 826–839.

Johnson-Laird, P. (2006) *How We Reason*, Oxford: Oxford University Press.

Jordan, K.E., Clark, K. & Mitroff, S.R. (2010) See an object, hear an object file: Object correspondence transcends sensory modality, *Visual Cognition*, **18** (4), pp. 492–503.

Kahneman, D., Treisman, A. & Gibbs, B.J. (1992) The reviewing of object files: Object-specific integration of information, *Cognitive Psychology*, **24**, pp. 175–219.

Kosslyn, S.M. (1980) *Image and Mind*, Cambridge, MA: Harvard University Press.

Kosslyn, S.M. (1994) *Image and Brain*, Cambridge, MA: MIT Press.

Neisser, U. (1967) *Cognitive Psychology*, Englewood Cliffs, NJ: Prentice-Hall.

Pearson, J., Naselaris, T., Holmes, E.A. & Kosslyn, S.M. (2015) Mental imagery: Functional mechanisms and clinical applications, *Trends in Cognitive Sciences*, **19** (10), pp. 590–602.

Pylyshyn, Z. (2003) *Seeing and Visualizing: It's Not What You Think*, Cambridge, MA: MIT Press.

Pylyshyn, Z. (2008) The empirical case for bare demonstratives in vision, in Viger, C. & Stainton, R.J. (eds.) *Compositionality, Context, and Semantic Values: Essays in Honour of Ernie Lepore*, pp. 255–274, New York: Springer.

Pylyshyn, Z. & Storm, R. (1988) Tracking multiple independent targets: Evidence for a parallel tracking mechanism, *Spatial Vision*, **3** (3), pp. 179–197.

Quilty-Dunn, J. (unpublished) *Syntax and Semantics of Perceptual Representation*, dissertation.

Scholl, B.J., Pylyshyn, Z. & Franconeri, S. (1999) When are spatiotemporal and featural properties encoded as a result of attentional allocation?, *Investigative Ophthalmology and Visual Science*, **40** (4), S797.

Scholl, B.J. & Leslie, A.M. (1999) Explaining the infant's object concept: Beyond the perception/cognition dichotomy, in Lepore, E. & Pylyshyn, Z. (eds.) *What is Cognitive Science?*, pp. 26–73, Oxford: Blackwell.

Sperling, G. (1960) The information available in brief visual presentations, *Psychological Monographs: General and Applied*, **74** (11), pp. 1–29.