

## Biederman and Cooper's 1991 paper

[Biederman I, Cooper E E, 1991 "Evidence for complete translational and reflectional invariance in visual object priming" *Perception* **20** 585–593. Original paper reprinted in the appendix.]

### Authors' update

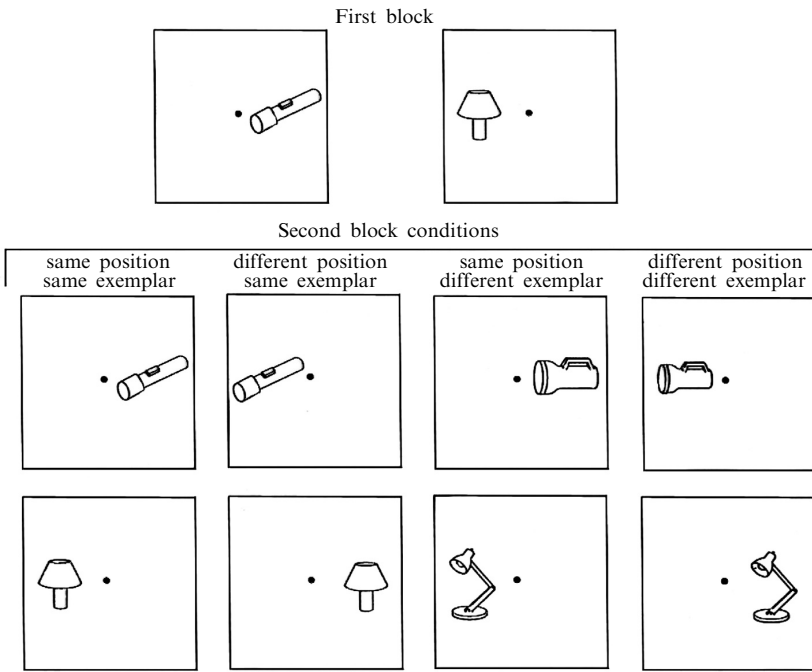
#### Translational and reflectional priming invariance: a retrospective

**Abstract.** Biederman and Cooper (1991a) showed that the presentation of a briefly presented image of an object at one position in the visual field facilitated its identification, as assessed by naming speed and accuracy, several minutes later. The facilitation was unaffected by a translation or a reflection of the stimulus. A component of this priming was visual rather than basic-level conceptual or lexical in that there was less facilitation for an object with the same name (and basic-level class) but a different shape. The invariance of priming to view variables has stood up well over the years and appears to be a general phenomenon—as long as the original structural description can be readily resolved—in that it has also been observed for variations in size and orientation in depth. Although priming was unaffected by a change in position, we documented that there was explicit memory for the position (and orientation and size) of the stimulus. The existence of two forms of representation from the identical stimulus presentation—one invariant and the other dependent on view variables—poses a challenge as to what can be concluded about view invariance from single-unit activity.

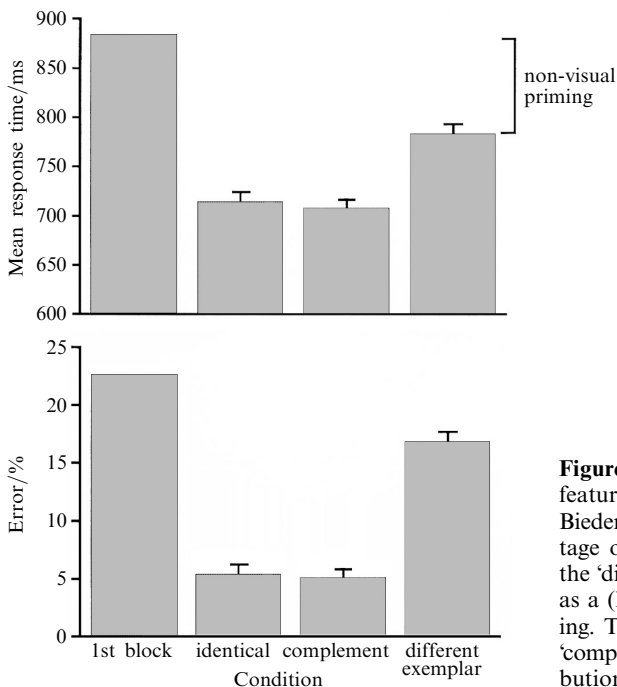
#### Introduction

In 1991 Biederman and Cooper (1991a) reported an experiment in which subjects named briefly presented, masked pictures (line drawings) of objects and then, in a second block about 7 min later, they would name the identical picture or a same name, different shaped exemplar (eg a grand piano followed by an upright piano). The pictures were either in the same or different positions and had the same or different left–right mirror orientations (figure 1). There was marked facilitation on the second block compared to the first for the identical pictures and this facilitation was greater for the identical pictures compared to the different exemplars (which also showed some facilitation). We interpreted the advantage of the identical images compared to the different exemplar images to be evidence of visual priming and the facilitation of the different exemplars as composed of a combination of non-visual priming, ie lexical and basic-level concept priming, and general practice in the task. (Because the different exemplars were physically more similar in shape to the original object than randomly selected objects, our partitioning could actually be regarded as a lower-bounds estimate of the visual priming component and, therefore, an upper-bounds estimate of the nonvisual priming and general practice components. Also, with high-quality stimuli, the magnitude of visual priming would be expected to be modest. With degraded stimuli, eg contour-deleted line drawings, the contribution of a prior exposure is considerably greater, as indicated in figure 2.)

The major contribution of the paper was in showing that the priming was unaffected by a left–right translation of 8.8 deg of an object. (The maximum extent of the objects was 4 deg and the nearest contour to fixation was 2.4 deg so the 8.8 deg translation was with respect to the objects' centers.) The translation invariance also held for vertical displacements, above or below fixation. Moreover, there was also invariance to the mirror reflection of the object. It is important to note that not only was there facilitation to translated and/or reflected images but there was virtually no cost as a result of the translation or the reflection in that such images were as quickly and as accurately named as those in their original position and orientation. These effects were replicated by Fiser and Biederman (2001) with grey-level images. DiCarlo and Maunsell (2003) reported that strong behavioral invariance to translation was evident in the object-matching performance of monkeys.



**Figure 1.** Illustration of conditions for the study of translation invariance (Biederman and Cooper 1991a). In name priming, subjects would simply name the objects, with any advantage of the first (leftmost) column, the identical condition, over the different exemplar condition (third column), a measure of visual priming. In an episodic recognition task, subjects experience the same first and second block stimuli as subjects in a priming experiment. Like subjects in a priming experiment, Cooper et al's (1992) episodic recognition paradigm required subjects to name the stimuli on the first block. On the second block, they had to judge whether the specific shape was the same as that on the first block, so correct performance would be to respond "old" for the stimuli in the two left columns and "new" to the stimuli in the two rightmost columns.



**Figure 2.** Results of the complementary feature-deleted experiment [adapted from Biederman and Cooper (1991b)]. The advantage of performance in the 'identical' over the 'different' exemplar condition is defined as a (lower-bound) estimate of visual priming. The equivalence of the 'identical' and 'complementary' conditions rules out a contribution of local features to the priming.

---

**Priming versus episodic (explicit) recognition**

The translation invariance that was so evident with name priming did not mean that subjects did not retain information about the object's position or orientation. Another group of subjects had the identical first block as the subjects in the priming experiment but rather than name the objects again in the second block, they judged whether the same objects were in the same or different position and/or orientation as in the first block. Subjects were well above chance in this task. In a refinement of this episodic recognition test, Cooper et al (1992) presented the identical stimuli on the second block as were shown in the priming tasks but now the subjects judged whether the objects were of the same shape ('old') or different shape ('new') as those shown in the first block, ignoring whether there was a change in position or orientation. The different exemplars constituted objects with a different shape (or 'new'). (The subjects were instructed that the new objects would have the same name but be of a different shape.) When an identical-shaped object was shown in a different position or orientation, there was a marked increase in error rates and RTs in judging that that object was shown on the first block. There is general acceptance that old–new judgments are largely based on feelings of familiarity and the change in position or orientation presumably lowered the familiarity of the representation. The interference from a change in view variables did not extend to the different exemplars, in that their judgments as 'new' items were unaffected by whether their different-shaped counterpart was shown in the same or different position or orientation. This latter result indicates that the episodic representation specified the specific (subordinate-level) shape that was shown on the first block, not just the name or basic-level concept. Priming and episodic representations were thus both dependent on the specific shape. The difference was that the priming was view-invariant.

**Other priming invariances**

The Biederman and Cooper (1991a) experiment was the first of several showing priming invariance and episodic dependence for view variables. Biederman and Cooper (1992) followed the (1991a) translation and reflection study with one that documented priming invariance for images that were of a different size (smaller or larger) than those viewed on the first block. In both the translation and size studies, the object images were line drawings. Fiser and Biederman (1995) extended the evidence for size invariance to gray-level images. Biederman and Gerhardstein (1993) employed naming to investigate the priming and matching of objects rotated in depth. As with translation, reflection, and changes in size, there was more priming for rotated identical pictures than different exemplars. However, according to geon theory (Biederman 1987), if a rotation of an object changes the visible parts and/or relations, ie the geon structural description changes, then a cost of recognition should be observed. [A structural description is a representation that specifies separate types of units for parts and relations (Humphreys and Riddoch 1987; Winston 1975).] This expectation was confirmed in direct comparisons of rotations that did or did not produce differences in the parts that were in view. See Biederman (2000) for a review of the effects on object recognition of rotation in depth.

**Methodological problems with same – different matching tasks**

In contrast to the longer-term priming tasks, there have been several reports of modest costs of translation (and changes in size or orientation) in same–different shape matching tasks (eg Dill and Edelman 2001; Ellis and Allport 1986) with brief (eg 1 s) interstimulus intervals (ISIs). However, as noted by Biederman and Bar (1999), these costs may be largely artifactual. On a 'same (shape)' trial in which the object is in its original position, there is no stimulus change between S1 (the first stimulus) and S2 (the second stimulus) so the subject can respond "same". However, on 'same

(shape) trials in which there is a translation, the subject had to determine whether the difference is caused by a change of shape (requiring a “different” response) or only a change of position (or size or orientation, requiring a “same” response), resulting in an elevation of RTs and error rates. This effect can occur even with an intervening mask if the subject, quite reasonably, bases his or her decision only on something that resembles S1. Biederman and Bar (1997) documented this artifact when they showed that apparent costs of depth rotation in same–different sequential shape matching tasks could be virtually eliminated by introducing a translation between S1 and S2. The effect of the translation was to elevate RTs and error rates on the 0° (non-rotated) ‘same’ trials, so that they were almost equal to those of the 30° and 60° rotation conditions. These later conditions were unaffected by the translation. Thus the apparent costs of rotation—the reported faster RTs and error rates at 0° rotation—were likely produced by a decision strategy that exploited the absence of any stimulus change in the 0° ‘same’ condition. If one wanted to investigate translation effects themselves in a same–different matching task, then some way of defeating a fast ‘same’ strategy for non-translated stimuli, such as a change in size on all trials, was required. An additional potentially biasing factor is that, unlike the longer-term name-priming paradigm, subjects can have a bias to either look at or think about the object at its original position. Even with these potential biasing factors, a number of researchers have reported translation invariance with highly similar, unfamiliar shapes (eg Dill and Edelman 2001, in three of five experiments; Bricolo and Bühlhoff 1992). These cases, where there was a failure to find complete translation invariance, have been difficult discriminations where the translation placed different parts at different retinal eccentricities, as discussed below in the section on priming contour-deleted images.

Another decision strategy compromising sequential same–different matching paradigms was revealed in a series of experiments by Dill and Edelman (2001) in their investigation of translation costs. These authors noted the practice in many same–different studies to report only the positive (same) trials (eg Hayward and Tarr 1997; Tarr et al 1998). Dill and Edelman found that these trials show a cost (higher error rates) compared to trials in which the object was translated. However, when they looked at the ‘different’ trials, they found the opposite effect, with higher error rates for objects that remained in their original position. So when ‘same’ and ‘different’ trials were averaged, there was no effect of translation!

The general lesson here is that, because same–different matching tasks invite strategies because of their low uncertainty about what has to be judged, one must be cautious about inferences made on the basis of such tasks. This does not mean that a naming task is not without problems. Consider a task in which a subject learned nonsense-syllable names to unfamiliar objects that, during training, were shown only at a single position or orientation (eg as in Tarr et al 1998). If there is a cost on test trials when the object is in a new position, would this mean that there is a translation cost, per se? Not necessarily. It could be that what is learned is an association of the name to a particular shape in a particular position. By failing to consider these caveats about same–different strategies, episodic memory, selection of only positive trials, and position-dependent response learning, it is likely that the conclusions questioning the existence of translation invariance in a recent review by Kravitz et al (2007) themselves should be questioned. Basic-level naming has an advantage in that the names have been learned independently of the position of the object and can be run under conditions of high image uncertainty so as to eliminate these potential confounding factors. It would be highly desirable if an independent method could be devised that could provide converging tests of invariance (see Biederman and Gerhardstein 1995, and Biederman and Bar 1999 for additional discussion of methodological issues).

---

### **What mediates priming?**

Can we describe the representation that mediates visual priming as assessed by naming performance of briefly presented images? Basic-level concept priming as well as lexical (word) priming is ruled out because subjects (presumably) activate the same basic-level concept and name for both identical and different-exemplar images. Also, having the subjects read the names of the objects prior to the experimental trials has no effect on picture-naming performance (Biederman 1987; Biederman and Cooper 1991a).

The remaining alternatives would appear to be (a) local features, such as the contours and vertices of the object, (b) the simple parts, or (c) a subordinate-level concept of the object. Biederman and Cooper (1991b) tested among these alternatives by deleting every other edge and vertex from one member of a complementary pair of line drawings. The other member was composed of the deleted features from the first member. So each member contained half the contour with a completely different set of vertices and line segments. From either member, the same simple parts could be activated (or not activated, if the deletion rendered the part unidentifiable). After seeing one contour-deleted image on the first block for a given object, a subject would see either the identical, complementary, or contour-deleted image of a same-name different exemplar on the second block. There was considerable visual priming in that the identical images on the second block were named markedly faster and more accurately than the different exemplars, but performance with the complementary images was equivalent to that with the identical images (figure 2). This result rules out a contribution of the local features (and spatial-frequency content) to the priming.

The first experiment ruled out a role of basic-level concept activation and lexical (name) access because performance on the identical condition was superior to that on the different exemplar condition (defining visual priming). A second experiment was required to rule out a contribution of subordinate-level priming. Although subjects might say "piano" when seeing either a grand piano or an upright piano, the representation mediating the priming might be the particular subordinate-level concepts. In this experiment, half the parts were deleted from each member of a complementary pair, but the parts themselves were intact. So each member of a complementary pair was of the same subordinate-level concept, eg a grand piano, with no parts in common. Now, there was no visual priming of the complement in that performance with the complement was identical to that of the different exemplars, eg an upright piano. This experiment thus ruled out a role of subordinate-level activation and established that there would be no priming if the images did not share the same parts. A possible concern was that members of a complementary pair of the local contour-deleted images in the first experiment were more similar to each other than the part-deleted members of a complementary pair in the second experiment so that overlap with a template might predict the greater complementary priming in experiment 1, although not the complete equivalence of the priming of complements and identical images. Biederman and Cooper (1991b) correlated the similarity in aspect ratios of members of parts-deleted complementary pairs with the magnitude of priming. The correlation was essentially zero. Still another finding inconsistent with a template account was the invariance in priming of mirror reflection, reported by Biederman and Cooper (1991a, 1991b) and Fiser and Biederman (2001) (see Fiser et al 1996, for additional controls).

### **Translation costs with contour-deleted images: Effects of eccentricity**

Using the local feature-deleted stimuli of Biederman and Cooper (1991b), in which every other vertex and line was deleted from each geon, Cooper et al (1992) reported reduced priming for a translation between 2.4 deg to the left or right of fixation of an image that subtended approximately 4 deg. The explanation offered by these authors

was that, given the greater difficulty of determining the parts of the object for these contour-deleted images, the translation brought different parts into the higher resolution region of the visual field—2.4 deg versus 6.4 deg. That is, if the nose of an airplane were closest to the fixation point when the object was on the right, the tail of the plane would be closest when the object was on the left. Consistent with this explanation was their finding that the translation costs were eliminated when the objects were not only translated but reflected as well. The reflection would have brought the same parts, but in a different left–right orientation, into the higher resolution region.

### **Neural basis of view-invariance**

If we assume that the inferior temporal (IT) cortex in the macaque provides the neural substrate mediating both priming and episodic recognition, with the former being position-invariant and the latter being position-dependent, then it is difficult to know just what to expect when investigating the neural effects of position changes.

### **Single-unit recording in the macaque**

Most studies of single-unit recording in IT in the macaque have documented large receptive fields (RFs) (10–30 deg) that almost always include the fovea (eg Boussaoud et al 1991; see DiCarlo and Maunsell 2003 for a recent review). Insofar as the response falls off with distance from the RF center, a strong form of neural translation invariance, in which a response to a stimulus would not vary with position within the RF, is not supported by the tuning data (Boussaoud et al 1991). However, Sary et al (1993) reported that, despite this variation in absolute firing rates within the RF, the preference for a particular stimulus was relatively constant.

One obvious reason why activity would vary with RF position is that retinal eccentricity is confounded with position, although this might not explain all of the tuning variation. An exception to the general finding of large RFs and preserved tuning preferences are the results of a study by DiCarlo and Maunsell (2003) who reported much smaller receptive fields with much less preference consistency with position variation than previous studies. However, as these authors note, their stimuli were markedly smaller and more similar than those used in other studies. Relevant here is Rolls and Deco's (2002) report that a cluttered background served to shrink the extent of the receptive field without a reduction in the magnitude of IT neuronal firing.

It is clear that IT neurons (can) convey information about position as well as shape. However, as noted by DiCarlo and Maunsell (2003), a statement of the relative sensitivity to the two kinds of information is problematic because it is likely a function of the particular stimuli selected and the range of variation. Whereas Tovée et al (1994) suggested that the information conveyed by IT neurons for faces was about four times that for position, DiCarlo and Maunsell (2003) reported about equal sensitivity.

With respect to reflection invariance, Logothetis and Pauls (1995) found that some IT cells responded to both a trained view of an object and its mirror reflection.

What is unknown is whether and how the information in the tuning for shape and position is ultimately read by neurons later in the processing stream. The problem of a neuron conveying information about more than one dimension is, of course, not unique to IT. V1 neurons simultaneously convey information about scale and orientation, yet an observer can readily respond to one of these dimensions while easily ignoring the other. The greatest obstacle in inferring function from the size of IT receptive fields, however, may be that different representations may be created from a single encounter with a stimulus (as described previously): an invariant one for shape as indexed by priming and an episodic memory that binds all aspects of the encounter (including view variables) with that shape. Small receptive fields in IT may be mediating episodic memory for position rather than position invariance.

---

## Human fMRI

There is some consensus that the human homologue to macaque IT includes a region in the occipital-temporal cortex termed the lateral occipital complex (LOC) (Orban et al 2004). This region is defined by greater activation of an intact object, familiar or unfamiliar, compared to a scrambled version of the same object (thus controlling for contrast energy). LOC is bilaterally activated and appears to be on a critical path for object recognition in that bilateral lesions to this region result in a complete loss of the ability to perceive shape for purposes of recognition (James et al 2003).

Invariances of the representations in LOC have been investigated with fMRI-A (A for adaptation), in which presentation of a repeated stimulus results in a reduced BOLD response compared to presentation of a novel stimulus, as a measure of priming. In general, experiments employing this method show that LOC is responsive to perceived shape rather than to the specific images. Grill-Spector et al (2001) showed that line drawings and grey-level photographs produced mutual fMRI-A priming equivalent to that produced by repetition of the identical image. Kourtzi and Kanwisher (2001) reported equivalent priming to an image of an object with bars either in front (and therefore partially occluding it) or behind it. These authors demonstrated that with figure-ground reversible images, it was the perceived shape that determined priming, despite presentation of identical images. If what was ground in the first presentation was the perceived shape in the second, there was no priming. Given these results, one might expect that the complementary local feature-deleted stimuli of Biederman and Cooper (1991b) would show similar fMRI-A priming and that parts-deleted complements would show no adaptation between members of a complementary pair. This is exactly what Hayworth and Biederman (2006) reported.

In one of the few studies of viewpoint effects on fMRI-A, Vuilleumier et al (2002) reported that variations in size or viewpoint (45° rotation in depth) had no effect on the magnitude of fMRI-A in the left fusiform gyrus, but a change in size or viewpoint did reduce fMRI-A in the right fusiform. The rotation could have changed parts but it is also possible, given the perspective of our position, that these effects might reflect priming and episodic memory, although other investigators have not reported lateralised fMRI-A priming effects in the fusiform.

Although there is strong evidence that object perception is invariant to absolute position, higher-level visual representations are extremely sensitive to relative position. Given a minimal novel scene of, say, an elephant above a bus, a translation of 3.3° results in no release from adaptation in LOC whereas a strong release is evident when switching the relative positions an equal extent so that the bus is now above the elephant (Hayworth et al 2008).

## Conclusion

The demonstration of translation and reflectance invariance by Biederman and Cooper (1991a) was only the first of a number of studies documenting that visual priming is mediated by a representation of the shape of an object largely independent of position, size, orientation in depth, reflection, and surface properties. At the same time, there is also an episodic representation specifying the original experience. There have been minimal challenges to these observations of long-term priming invariance and episodic dependence of view variables. In general, the representation mediating priming reflects the shape as it is perceived, rather than the image from which that percept is extracted. Both forms of representation are highly useful but this dual representation from a single presentation poses a challenge as to what can be concluded about view invariance from single-unit activity.

**Acknowledgments.** We thank Kenneth J Hayworth for his helpful comments. This research was supported by NSF BCS-0420794, 0531177, and 0617699 to IB.

Irving Biederman

Department of Psychology, University of Southern California, Los Angeles, CA 90089-2520, USA;  
e-mail: bieder@usc.edu

Eric E Cooper

Department of Psychology, Iowa State University, Ames, IA 50011, USA

## References

- Biederman I, 1987 "Recognition-by-components: A theory of human image understanding" *Psychological Review* **94** 115–147
- Biederman I, 2000 "Recognizing depth-rotated objects: A review of recent research and theory" *Spatial Vision* **13** 241–253
- Biederman I, Bar M, 1997 "What's the fuss about perceiving depth-rotated objects?" Paper presented at the Meetings of the Psychonomics Society, Philadelphia, PA, November
- Biederman I, Bar M, 1998 "Same–different matching of depth-rotated objects" *Investigative Ophthalmology & Visual Science* **39** 1113 (abstract)
- Biederman I, Bar M, 1999 "One-shot viewpoint invariance in matching novel objects" *Vision Research* **39** 2885–2899
- Biederman I, Cooper E E, 1991a "Evidence for complete translational and reflectional invariance in visual object priming" *Perception* **20** 585–593
- Biederman I, Cooper E E, 1991b "Priming contour-deleted images: Evidence for intermediate representations in visual object recognition" *Cognitive Psychology* **23** 393–419
- Biederman I, Cooper E E, 1992 "Size invariance in visual object priming" *Journal of Experimental Psychology: Human Perception and Performance* **18** 121–133
- Biederman I, Gerhardstein P C, 1993 "Recognizing depth-rotated objects: Evidence and conditions for 3D viewpoint invariance" *Journal of Experimental Psychology: Human Perception and Performance* **19** 1162–1182
- Biederman I, Gerhardstein P C, 1995 "Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff" *Journal of Experimental Psychology: Human Perception and Performance* **21** 1506–1514
- Boussaoud D, Desimone R, Ungerleider L, 1991 "Visual topography of area TEO in the macaque" *Journal of Comparative Neurology* **306** 554–575
- Bricolo E, Bülthoff H H, 1992 "Translation-invariant features for object recognition" *Perception* **21** Supplement 2, 59 (abstract)
- Cooper E E, Biederman I, Hummel J E, 1992 "Metric invariance in object recognition: A review and further evidence" *Canadian Journal of Psychology* **46** 191–214
- DiCarlo J J, Maunsell J H R, 2003 "Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position" *Journal of Neurophysiology* **89** 3264–3278
- Dill M, Edelman S, 2001 "Imperfect invariance to object translation in the discrimination of complex shapes" *Perception* **30** 707–724
- Ellis R, Allport D A, 1986 "Multiple levels of representation for visual objects; A behavioural study", in *Artificial Intelligence and Its Applications* Eds A G Cohen, J R Thomas (New York: John Wiley) pp 245–257
- Fiser J, Biederman I, 1995 "Size invariance in visual object priming of gray-scale images" *Perception* **24** 741–748
- Fiser J, Biederman I, 2001 "Invariance of long-term visual priming to scale, reflection, translation, and hemisphere" *Vision Research* **41** 221–234
- Fiser J, Biederman I, Cooper E E, 1996 "To what extent can matching algorithms based on direct outputs of spatial filters account for human shape recognition?" *Spatial Vision* **10** 237–271
- Grill-Spector K, Kourtzi Z, Kanwisher N, 2001 "The lateral occipital complex and its role in object recognition" *Vision Research* **41** 1409–1422
- Hayward W G, 1998 "Effects of outline shape in object recognition" *Journal of Experimental Psychology: Human Perception and Performance* **24** 427–440
- Hayward W G, Tarr M J, 1997 "Testing conditions for viewpoint invariance in object recognition" *Journal of Experimental Psychology: Human Perception and Performance* **23** 1511–1521
- Hayworth K J, Biederman I, 2006 "Neural evidence for intermediate representations in object recognition" *Vision Research* **46** 4024–4031
- Hayworth K, Lescroart M, Biederman I, 2008 "Explicit relation coding in the lateral occipital complex" <http://journalofvision.org/8/6/35/>



- 
- Humphreys G W, Riddoch M J, 1987 *Visual Object Processing: A Cognitive Neuropsychological Approach* (London: Lawrence Erlbaum Associates)
- James T W, Culham J, Humphrey G K, Milner A D, Goodale M A, 2003 "Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study" *Brain* **126** 2463–2475
- Kayaert G, Biederman I, Vogels R, 2003 "Shape tuning in macaque inferior temporal cortex" *Journal of Neuroscience* **23** 3016–3027
- Kourtzi Z, Kanwisher N, 2001 "Representation of perceived object shape by the human lateral occipital cortex" *Science* **293** 1506–1509
- Kravitz D J, Vinson L D, Baker C I, 2008 "How position-dependent is visual object recognition?" *Trends in Cognitive Sciences* **12** 114–122
- Logothetis N K, Pauls J, 1995 "Psychophysical and physiological evidence for view-centered object representations in the primate" *Cerebral Cortex* **5** 270–288
- Orban G A, VanEssen D, Vanduffel W, 2004 "Comparative mapping of higher visual areas in monkeys and humans" *Trends in Cognitive Sciences* **8** 315–324
- Rolls E T, Deco G, 2002 *The Computational Neuroscience of Vision* (Oxford: Oxford University Press)
- Sary G, Vogels R, Orban G A, 1993 "Cue-invariant shape selectivity of macaque inferior temporal neurons" *Science* **260** 995–997
- Tarr M J, Williams P, Hayward W G, Gauthier I, 1998 "Three-dimensional object recognition is viewpoint dependent" *Nature Neuroscience* **1** 275–277
- Tovée M J, Rolls E T, Azzopardi P, 1994 "Translation-invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque" *Journal of Neurophysiology* **72** 1049–1060
- Vuilleumier P, Henson R N, Driver J, Dolan R J, 2002 "Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming" *Nature Neuroscience* **5** 491–499
- Winston P H, 1975 "Learning structural descriptions from examples", in *The Psychology of Computer Vision* Ed. P H Winston (New York: McGraw-Hill)

---

## Comments

### Challenges in object recognition: Selectivity vs invariance

Successful interactions in the complex environments we inhabit entail that we recognise objects independently of local image changes in their properties (eg position, size, illumination, orientation) that occur as observers and objects dynamically interact and move in natural scenes. Despite the ease and speed with which we recognise objects from different categories, this is a computationally demanding task for the visual system. The complexity of this challenge is augmented by the fact that information about these spatial image properties is critical for navigation and interactions with objects. If this information was to be discarded for the sake of invariant object recognition, it would be at the cost of successful actions. How does the brain manage to strike the right balance between invariant object representations and feature selectivity to support both recognition and actions in dynamic environments?

Biederman and Cooper in 1991 provided the first insights into answering this question. In an elegant study comparing priming and recognition tasks, they provide evidence for explicit memory of spatial relationships (object position, left–right orientation) while implicit representations of object identity are invariant to changes in these image properties. This work was followed by numerous studies investigating recognition across different types of image changes (eg rotations in the image plane or in depth) and inspired a controversial debate on the nature of visual object representations. In particular, structural-description models suggest that objects are represented on the basis of the configuration of volumetric primitive parts that is not affected by image transformations (Biederman 1987; Hummel and Biederman 1992; Marr and Nishihara 1978). On the contrary, image-based models propose that objects are represented as sets of multiple 2-D views and transformed object images can be recognised by extrapolating or interpolating between these known object views (Bülthoff and Edelman 1992; Bülthoff et al 1995; Edelman 1995).

Despite the apparent controversy between these theoretical approaches, they both point to the same fundamental question regarding the origins of object representations in the visual system. How does the brain construct object representations? How do neurons become tuned to global feature configurations and selective to subtle changes in object identity, while remaining tolerant to image changes that preserve object identity? It has been suggested that experience and learning play an important role in the optimisation of the visual system. Encoding of invariant primitives may relate to evolutionary and developmental plasticity mechanisms for shaping object selectivity and invariance by taking into account statistical regularities in the environment (eg collinearity, circularity, symmetry). At a shorter time scale, experience through everyday interactions with objects and exposure to dynamic object transformations may shape the tuning of neural populations in a probabilistic manner and result in selectivity and invariance for feature configurations that co-occur frequently.

Recent neurophysiological and neuroimaging studies have investigated the neural architecture that supports robust object recognition and the role of experience in shaping selectivity for objects and tolerance to image changes. Specifically, while local image features (position, orientation) are known to be processed in primary visual cortex, complex shapes are represented towards the end of the visual pathway in the temporal cortex (Felleman and Van Essen 1991; Grill-Spector and Malach 2004). In particular, neural populations in higher temporal areas contain information about object identity that generalises across image changes (eg Grill-Spector and Malach 2004; Hung et al 2005; Quiroga et al 2005; Rolls 2000). Computational models (Fukushima 1980; Riesenhuber and Poggio 1999; Ullman and Soloviev 1999) propose that the brain builds these robust object representations using neuronal connections that group together

similar image features across image transformations. Recent neurophysiological studies (Zoccolan et al 2007), showing that temporal cortex neurons with high object selectivity have low invariance, suggest that such connections between neurons selective for similar features are critical for the binding of feature configurations and the robust representation of object identity. These findings suggest that while individual neurons contain highly selective information for image features that is critical for interactions with objects (eg precise information about position, size, orientation), connections across neural populations may support object recognition across image changes.

But how does the brain know which neurons to connect or which connections across neural populations to strengthen to build robust object representations? Experience with the statistical regularities in natural scenes has been proposed as a potential solution to this problem (Foldiak 1991; Ullman and Soloviev 1999; Wallis and Bülthoff 2001; Wallis and Rolls 1997; Wiskott and Sejnowski 2002). In particular, the brain capitalises on the spatiotemporal statistics of the environment and learns to bind image features that co-occur frequently into global configurations. For example, it has been suggested that temporal continuity and motion enhance the binding of disparate images into the same object representation (Cox et al 2005; Kourtzi and Shiffrar 1997; Wallis and Bülthoff 2001). At the neural level, neurons in the temporal cortex show enhanced selectivity after training for novel objects and generalisation across neighbouring views to the trained orientation (Kobatake et al 1998; Logothetis et al 1995; Rolls 1995). Such learning-dependent plasticity mechanisms in the temporal cortex may support robust object recognition across images as proposed by Biederman and Cooper. Finally, it is possible that similar mechanisms may mediate long-term optimisation through evolution and development and tune the visual system to the fundamental principles of feature binding and the shape primitives critical for object recognition.

Zoe Kourtzi

School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK;  
e-mail: z.kourtzi@bham.ac.uk

## References

- Biederman I, 1987 "Recognition-by-components: a theory of human image understanding" *Psychological Review* **94** 115–147
- Bülthoff H H, Edelman S, 1992 "Psychophysical support for a two-dimensional view interpolation theory of object recognition" *Proceedings of the National Academy of Sciences of the USA* **89** 60–64
- Bülthoff H H, Edelman S Y, Tarr M J, 1995 "How are three-dimensional objects represented in the brain?" *Cerebral Cortex* **5** 247–260
- Cox D D, Meier P, Oertelt N, DiCarlo J J, 2005 "'Breaking' position-invariant object recognition" *Nature Neuroscience* **8** 1145–1147
- Edelman S, 1995 "Representation of similarity in three-dimensional object discrimination" *Neural Computation* **7** 408–423
- Felleman D J, Van Essen D C, 1991 "Distributed hierarchical processing in the primate cerebral cortex" *Cerebral Cortex* **1** 1–47
- Foldiak P, 1991 "Learning invariance from transformation sequences" *Neural Computation* **3** 194–200
- Fukushima K, 1980 "Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position" *Biological Cybernetics* **36** 193–202
- Grill-Spector K, Malach R, 2004 "The human visual cortex" *Annual Review of Neuroscience* **27** 649–677
- Hummel J E, Biederman I, 1992 "Dynamic binding in a neural network for shape recognition" *Psychological Review* **99** 480–517
- Hung C P, Kreiman G, Poggio T, DiCarlo J J, 2005 "Fast readout of object identity from macaque inferior temporal cortex" *Science* **310** 863–866
- Kobatake E, Wang G, Tanaka K, 1998 "Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys" *Journal of Neurophysiology* **80** 324–330
- Kourtzi Z, Shiffrar M, 1997 "One-shot view invariance in a moving world" *Psychological Science* **8** 461–466

- 
- Logothetis N K, Pauls J, Poggio T, 1995 "Shape representation in the inferior temporal cortex of monkeys" *Current Biology* **5** 552–563
- Marr D, Nishihara H K, 1978 "Representation and recognition of the spatial organization of three-dimensional shapes" *Proceedings of the Royal Society of London, Series B* **200** 269–294
- Quiroga R Q, Reddy L, Kreiman G, Koch C, Fried I, 2005 "Invariant visual representation by single neurons in the human brain" *Nature* **435** 1102–1107
- Riesenhuber M, Poggio T, 1999 "Hierarchical models of object recognition in cortex" *Nature Neuroscience* **2** 1019–1025
- Rolls E T, 1995 "Learning mechanisms in the temporal lobe visual cortex" *Behavioural Brain Research* **66** 177–185
- Rolls E T, 2000 "Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition" *Neuron* **27** 205–218
- Ullman S, Soloviev S, 1999 "Computation of pattern invariance in brain-like structures" *Neural Networks* **12** 1021–1036
- Wallis G, Bühlhoff H H, 2001 "Effects of temporal association on recognition memory" *Proceedings of the National Academy of Sciences of the USA* **98** 4800–4804
- Wallis G, Rolls E T, 1997 "Invariant face and object recognition in the visual system" *Progress in Neurobiology* **51** 167–194
- Wiskott L, Sejnowski T J, 2002 "Slow feature analysis: unsupervised learning of invariances" *Neural Computation* **14** 715–770
- Zoccolan D, Kouh M, Poggio T, DiCarlo J J, 2007 "Trade-off between object selectivity and tolerance in monkey inferotemporal cortex" *Journal of Neuroscience* **27** 12292–12307
- 

There is perhaps an art to writing papers that eventually become classics. Although we do not quite understand what the rules of this art-form are, a few characteristics would appear to be essential: (i) tackling a 'big' question; (ii) keeping the experimental methodology as simple as possible; (iii) having easily interpretable results; and (iv) presenting and defending a bold bottom-line conclusion. As I elaborate below, the paper by Biederman and Cooper embodies all of these attributes. In hindsight, it is not surprising that it became as influential as it did.

Pondering about which aspect of his and James Watson's work was most significant, Francis Crick wrote in his memoirs: "The major credit I think Jim and I deserve, ... is for selecting the right problem and sticking to it. It's true that by blundering about we stumbled on gold, but the fact remains that we were looking for gold". The 'gold' that Biederman and Cooper were looking for was an understanding of one of the most fundamental issues in visual cognition—what is the nature of internal object representations that subserve recognition? The primary reason this question is so challenging is that we hardly ever see an object the same way during different viewings. The object's image falls on different regions of the retina at different times, the lighting changes, and we view it from various directions and distances. Are our object representations versatile enough to tolerate these variations? If one had to pick just one key question in visual cognition, this would probably be it. Just by choosing this as the theme of their paper, Biederman and Cooper had already ensured that their results, irrespective of how they turned out, would be of interest and relevance to most researchers in the domain of high-level vision.

The experimental paradigm that Biederman and Cooper used to address this question was simplicity itself. Subjects were asked merely to name line drawings of commonplace objects and their reaction times were recorded. By adopting this paradigm, the authors rendered themselves immune to concerns regarding the results possibly being an artifact of complex instructions or tasks. Furthermore, the naming task mimicked natural recognition scenarios, and its results could thus be legitimately regarded as having relevance to real-world settings.

The data that this experimental paradigm yielded were impressive in how 'clean' they were. During the second viewing of a given image, all of the transformations tested were comparable to the untransformed image in the extent of priming they induced.

The authors could discern no statistically significant costs to changes in object position or viewpoint. This surprising result led them to a conclusion with far-reaching implications—the internal representations of objects for recognition are invariant to translation and rotation. I am sure Biederman and Cooper must have realised that this was an audacious message. But, I am impressed that they did not let worries about the debate that was sure to follow dilute their conclusion. The message is crisp. They had stated their position unambiguously and could now sit back and see how the world reacted to it.

The reaction was a spirited one. While several experiments that followed corroborated Biederman and Cooper's hypothesis, many experiments challenged the invariance idea by showing that even fairly small transformations in the appearance of an object could lead to discernible effects in recognition performance. I was a graduate student at the time when the debate was at its most vigorous. It made for interesting viewing. At several conferences, I and many of my colleagues got a first-hand look at a passionate back and forth amongst scientists who wanted to get at the bottom of this big question. What I admired about Irv was his ability to reconcile seemingly difficult data from other labs with the hypothesis he and Cooper had framed. He defended his ideas with logic and, always, with wit and gusto. These experiences made me realise how important a well-framed piece of research can be. In taking a step towards a grand truth, one can catalyze others to do so too. This, in essence, moves the entire field forward. And this is precisely what Biederman and Cooper managed to do with their 1991 paper.

Pawan Sinha

46-4077, Department of Brain and Cognitive Sciences, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; e-mail: psinha@mit.edu; website: <http://web.mit.edu/bcs/sinha/home.html>

### **Invariant parts of a citation classic**

Biederman and Cooper's *Perception* article on translational and reflectional invariance of visual object priming has become a widely cited paper<sup>(1)</sup> because it combined a number of important issues that were also quite hot at the time. First of all, the study reported in this paper addressed one of the core ideas of Biederman's (1987) highly influential recognition-by-components (RBC) theory, that the representations of shape needed for object recognition are invariant to changes in viewpoint. This particular set of experiments concerned invariance to translation and reflection, while other studies by Biederman and colleagues around the same time concerned invariance to changes in size (Biederman and Cooper 1992) and in depth orientation (Biederman and Gerhardstein 1993). Especially the notion of viewpoint-independence of object representations has proven very controversial, partly because behavioural evidence alone is often indecisive regarding invariance (Wagemans et al 1996) and partly because the mathematical notion of invariants under transformations (Van Gool et al 1994) has not always been taken seriously enough. Viewpoint-(in)dependence has been a critical issue in the debate on object-based versus image-based theories of object recognition, which dominated the field in the nineties. Because several later studies showed intermediate results (eg Foster and Gilson 2002; Vanrie et al 2001; Willems and Wagemans 2001), the controversy has now given room to compromise positions and fruitful empirical research investigating the conditions under which viewpoint-(in)dependence occurs (eg Demeyer et al 2007; Hayward 2003).

<sup>(1)</sup>At the time of the proofs of this article (May 20, 2009), ISI's Web of Science (SSCI edition) yielded 187 citations for Biederman and Cooper (1991a). Other papers from the same line of research were equally big hits (eg, Biederman and Cooper 1991b: 182; Biederman and Cooper 1992: 209; Biederman and Gerhardstein 1993: 315; Hummel and Biederman 1992: 390). The theoretical paper that has been the source of inspiration for this highly productive line of research (Biederman 1987) has been cited no less than 1711 times to date.

---

A second important aspect of Biederman's RBC theory was the notion that object representations consisted of parts, which could be modeled as primitive, geometric building blocks ('geons'). The beauty of this notion was that a restricted number of image regularities (eg collinearity, parallelism, symmetry), which could be considered as qualitative, non-accidental properties (eg Kukkonen et al 1996; Wagemans 1992, 1993; Wagemans et al 2000), was supposed to be sufficient to derive these 3-D building blocks quickly and reliably from bottom-up image cues. This idea enabled an important role for Gestalt grouping principles within a cognitive theory of object representation and recognition. Together with the computational Marr-like flavour of the theory [which was later implemented in a neural network by Hummel and Biederman (1992)], this synthesis paved the way to success amongst the widest possible audience of vision scientists and cognitive scientists alike. The issue of object segmentation into parts and the relevance of parts in object representations have since then attracted a great deal of attention, also outside the context of Biederman's RBC theory (eg Barenholtz and Feldman 2003; Bertamini and Mosca 2004; De Winter and Wagemans 2006).

The way in which Biederman and colleagues have tested the importance of part-based descriptions of object shape was by using contour-deleted pictures. In one of their studies using a priming paradigm (Biederman and Cooper 1991), they compared pictures of different exemplars to pictures of the same exemplars, yielding a nonvisual and a visual component of the priming effect (ie the advantage of naming the pictures in the second block of trials compared to the first block). Moreover, they also showed that complementary fragments worked as well as identical fragments, demonstrating that all of the visual priming was situated at the intermediate, part-based level, not the low, image-based level of contour fragments. Although the specific nature of contour-deletion may have contributed to these results (see Panis et al 2008), the results were clearcut and striking in their own right.

Part of the controversial nature of these experiments and the surrounding theories had to do with the use of the above long-term priming paradigm (subsequent prime and test blocks of trials rather than prime-target pairs of trials) and naming as a task (rather than same-different matching). As discussed in Biederman and Cooper's retrospective (see their update), these differences between experimental paradigms may explain the sometimes conflicting results in later studies, with some showing complete invariance to changes in view variables and others showing only partial invariance or lack of invariance. Such a pattern of conflicting results, along with the amount of discussion about the relevance of different paradigms that it has inspired, has definitely contributed to the citation levels of these studies.

A final charming aspect of the study—one that has also enabled contact with a different literature—was the use of an explicit (episodic) recognition task to demonstrate that participants could remember the original position of the pictures above chance levels. This established a dissociation between explicit recognition, which showed that position and orientation had been encoded, and implicit priming, which showed translational and reflectional invariance. This kind of dissociation between explicit and implicit memory was quite hot in the memory literature at that time (eg Tulving and Schacter 1990).

In summary, this paper has become a citation classic because it started from a very influential theory (synthesising successful ideas), it presented clear but controversial results, and it made contact with issues that were also relevant to other related fields.

Johan Wagemans

Laboratory of Experimental Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium;  
e-mail: johan.wagemans@psy.kuleuven.be; website: <http://ppw.kuleuven.be/labexppsy/johanw/wagemans.htm>

---

**References**

- Barenholtz E, Feldman J, 2003 "Visual comparisons within and between object parts: Evidence for a single-part superiority effect" *Vision Research* **43** 1655–1666
- Bertamini M, Mosca F, 2004 "Early computation of contour curvature and part structure: Evidence from holes" *Perception* **33** 35–48
- Biederman I, 1987 "Recognition-by-components: A theory of human image understanding" *Psychological Review* **94** 115–147
- Biederman I, 2008 "Translational and reflectional priming invariance: A retrospective" This volume
- Biederman I, Cooper E E, 1991 "Priming contour-deleted images: Evidence for intermediate representations in visual object recognition" *Cognitive Psychology* **23** 393–419
- Biederman I, Cooper E E, 1992 "Size invariance in visual object priming" *Journal of Experimental Psychology: Human Perception and Performance* **18** 121–133
- Biederman I, Gerhardstein P C, 1993 "Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance" *Journal of Experimental Psychology: Human Perception and Performance* **19** 1162–1182
- Demeyer M, Zaenen P, Wagemans J, 2007 "Low-level correlations between object properties and viewpoint can cause viewpoint-dependent object recognition" *Spatial Vision* **20** 79–106
- De Winter J, Wagemans J, 2006 "Segmentation of object outlines into parts: A large-scale, integrative study" *Cognition* **99** 275–325
- Foster D H, Gilson S J, 2002 "Recognizing novel three-dimensional objects by summing signals from parts and views" *Proceedings of the Royal Society of London, Series B* **269** 1939–1947
- Hayward W G, 2003 "After the viewpoint debate: Where next in object recognition?" *Trends in Cognitive Sciences* **7** 425–427
- Hummel J E, Biederman I, 1992 "Dynamic binding in a neural network for shape recognition" *Psychological Review* **99** 480–517
- Kukkonen H T, Foster D H, Wood J R, Wagemans J, Van Gool L, 1996 "Qualitative cues in the discrimination of affine-transformed minimal patterns" *Perception* **25** 195–206
- Panis S, De Winter J, Vandekerckhove J, Wagemans J, 2008 "Identification of everyday objects on the basis of fragmented versions of outlines" *Perception* **37** 271–289
- Tulving E, Schacter D L, 1990 "Priming and human memory systems" *Science* **247** 301–306
- Van Gool L, Moons T, Pauwels E, Wagemans J, 1994 "Invariance from the Euclidean geometer's perspective" *Perception* **23** 547–561
- Vanrie J, Willems B, Wagemans J, 2001 "Multiple routes to object matching from different viewpoints: Mental rotation versus invariant features" *Perception* **30** 1047–1056
- Wagemans J, 1992 "Perceptual use of nonaccidental properties" *Canadian Journal of Psychology* **46** 236–279
- Wagemans J, 1993 "Skewed symmetry: A nonaccidental property used to perceive visual forms" *Journal of Experimental Psychology: Human Perception and Performance* **19** 364–380
- Wagemans J, Van Gool L, Lamote C, 1996 "The visual system's measurement of invariants need not itself be invariant" *Psychological Science* **7** 232–236
- Wagemans J, Van Gool L, Lamote C, Foster D H, 2000 "Minimal information to determine affine shape equivalence" *Journal of Experimental Psychology: Human Perception and Performance* **26** 443–468
- Willems B, Wagemans J, 2001 "Matching multi-component objects from different viewpoints: Mental rotation as normalization?" *Journal of Experimental Psychology: Human Perception and Performance* **27** 1090–1115

---

## Authors' response

We are most pleased and grateful for the kind words of Drs Sinha, Kourtzi, and Wagemans on the contribution of our 1991 *Perception* article and the retrospective on translation and reflection invariance in visual object priming.

It is curious that all three commentaries referred to the issue of invariance as "controversial" (Kourtzi and Wagemans) or one that "led to a spirited debate" (Sinha, paraphrasing ours). We term it 'curious' in that the controversy refers not to whether object recognition is invariant to translation (and reflection)—the subjects of our original paper—but whether it is invariant to rotation in depth. (We anticipated that the issue of rotation would be raised in the commentaries, which is why we discussed it in our retrospective.) Translation invariance seems so apparent to naive perception—that objects are readily recognised when viewed at different positions—that the target piece might have been accused of belabouring the obvious. It becomes a bit of a mystery, however, how this can be achieved, so the finding of translation invariance without cost poses a significant challenge, even today, to theorists.

Rotation in depth, of course, can produce much more drastic retinal image changes and a claim of strong invariance to such rotation is more striking, although (again) not in the eyes of the naive realist in that the term 'object constancy' was coined a century ago to refer to the unchanging perception of the shape of an object despite the image changes as it was rotated in depth. But this should never have been an issue in the first place! Biederman's (1987) recognition-by-components (RBC) paper offered a componential recovery principle proposing that the similarity between the components (parts, or geons, in their specified relations) between two views of an object should be the basis of the prediction of transfer between the two images. A speculation in that discussion was that the same similarity function might hold for rotation in depth as well as part deletion (or accretion) and occlusion. To the extent that the representation, a geon structural description specifying a non-accidental characterisation of the parts and their relations according to RBC, of the image of an object changes from one view to the next, RBC has to predict a cost. For translation of an image of an object there should be no cost, other than what may occur from the effects of retinal eccentricity. And this is what has been found with the most defensible measure of visual priming: long-term picture naming.

The issue of rotation in depth has been phrased as one of invariance versus 'view-based'. This is a most unfortunate characterisation on several grounds. First, all visual perception is view based. View-based is only an interesting possibility when discussing ESP! Second, the empirical issue cannot be whether there is or is not a cost of rotation—as RBC also predicts costs when geons or their resolution changes, or as noted above—but the conditions under which costs might be witnessed (Biederman and Gerhardstein 1993, 1995). Third, understanding what happens to recognition when objects are rotated requires specification of a theory of a representation. At best, 'view-based' is an effect (a cost of rotation) rather than a theory. The alternative to RBC implied by 'view-based' is that of a template. But large costs from rotation in depth are observed when there is a change in parts (Biederman and Gerhardstein 1993; Hayward 1998) or when non-accidental properties are insufficient to distinguish among the distractors (Biederman and Bar 1999), both of which are predicted from RBC and neither predicted from template accounts that make neither parts nor nonaccidental properties (NAPs) explicit. As noted by Sinha, costs are sometimes observed from small rotation angles in same–different tasks. When artifacts are removed (as can occur from transients in a same–different task) and near-accidents of critical information avoided (Biederman and Gerhardstein 1993, 1995) as described in our retrospective, the costs of rotation in depth are miniscule, if not absent.



---

We agree with Kourtzi that correlations of image measures over space and time are the likely candidates for providing the statistics supporting the connectivity underlying higher-level vision. Indeed, such correlations may be the basis of the greater sensitivity to NAPs and simple parts that can allow one-shot view invariance of novel objects.

ISSN 0301-0066 (print)

ISSN 1468-4233 (electronic)

# PERCEPTION

VOLUME 38 2009

[www.perceptionweb.com](http://www.perceptionweb.com)

**Conditions of use.** This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.