PERGAMON

Vision Research 40 (2000) 3901–3905

Vision Research

www.elsevier.com/locate/visres

Differing views on views: response to Hayward and Tarr (2000)

Irving Biederman [a,*], Moshe Bar [b]

[a] *Department of Psychology and Neuroscience Program, University of Southern California, Hedco Neurosciences Building, MC 2520, Los Angeles, CA 90089-2520, USA*

[b] *Massachusetts General Hospital, NMR Center, Bldg 149, Charlestown, MA 02129, USA*

Received 31 March 2000; received in revised form 14 June 2000

At the outset we should emphasize that nothing in Hayward and Tarr's commentary speaks to the main results of the Biederman and Bar (1999) report: Slight costs of rotation in detecting geon differences when matching a sequential pair of novel objects (3.3% increase in error rates), but massive costs (46.2% increase in error rates) in the detection of metric differences. Their comments are all addressed to our review of the literature of the relatively small costs (versus zero costs) that have been observed when geon differences were, presumably, available. Although they characterize some of Biederman and Bar's arguments as 'incorrect', in fact these arguments can readily be defended. Given the sensitivity that prompted these authors to write their comments, it is disappointing that their own characterization of many of Biederman and Bar's points are inaccurate.

It is not completely clear why the origin of these relatively small rotation costs are an important theoretical issue to Hayward and Tarr insofar as they have now 'eschewed' mental rotation mechanisms (their note 2). Instead, viewpoint costs are now regarded by these authors as reflecting changes in information and that 'there is no directly causal relation between the magnitude of a change in viewpoint of an object, and the magnitude of the associated cost in recognition' (note 2). This has been our position all along (Biederman, 1987; Biederman & Gerhardstein, 1993, 1995; Biederman & Bar, 1999).

1. Why assigning arbitrary names to novel objects may be problematic for assessing the view dependence/independence of object representations

Consider a visual classification task in which brief masked pictures of tables are presented to the left of fixation and brief masked pictures of chairs are presented to the right of fixation. An observer required to distinguish between chairs and tables might readily use the left–right view information to achieve high accuracy on such a task. However, few would accept such evidence of 'view-based recognition' as addressing issues concerned with the 'mental representation of objects' especially if that mental representation was to be interpreted as a representation of shape. An alternative position, proposed by Biederman and Cooper (1992), is that these viewpoint effects may be part of the episodic representation that specifies view variables and situation variables bound to a representation, perhaps implicit, of shape. That there might be at least two representations of objects is evidenced by Biederman and Cooper's (1992) finding that changes in the size of an image had no effect on basic-level name priming but produced marked interference in old-new, episodic recognition judgments. It is important to note that the first block of trials — basic level naming — was identical for the two tasks (Unlike experimental trained naming of novel stimuli, basic level naming is well practiced to be independent of view variables, such as an object's position in the visual field or whether it is facing left or right). Biederman and Cooper's interpretation was that the basic-level name priming was mediated by a representation invariant to view variables but that the old–new judgments were mediated by an episodic representation that combined shape and view-variables. Similar results (no effects on priming but large costs on old judgements) were reported by Biederman and Cooper (1991a,b) and Cooper, Biederman, and Hummel (1992) for position, reflection, and orientation.

Hayward and Tarr's comments would imply that there is a single, all-purpose, representation and that any variable that affects object recognition performance is, ipso facto, reflecting this single representation. Our position is perhaps more consistent with the classical

* Corresponding author. Tel.: +1-213-7406094; fax: +1-213-7405687.

E-mail address: bieder@usc.edu (I. Biederman).

PII: S0042-6989(00)00180-2

phenomenon of object constancy — that there is a representation of an object's shape independent of its position, size, and orientation (up to occlusion and accretion). However, there is also an episodic representation that, in combining shape with view variables, might be employed on a particular object recognition task. The degree to which it is employed, as noted by Biederman and Gerhardstein (1993), might well depend on its scale. Large shape differences would dominate smaller scale view variables but large view differences could dominate small shape differences, as suggested by the tables on left, chairs on right example.

In any event, as argued by Biederman and Gerhardstein (1993), the absence of rotation costs means that there is little to explain, assuming that there is sufficient power in the design to have detected a cost. The presence of rotation costs obligates the investigator to determine the cause of such costs. Hayward and Tarr's point that the rotation costs with naming tasks are consistent with other rotation costs is not particularly persuasive insofar as the origin of these costs is not clear. In addition to reflecting episodic representations, these costs may stem from resolution difficulties and transient shifts, as described below. Most important, given the immediate and extraordinarily large gain from distinctive GSDs, which are almost always available when humans are attempting to distinguish basic- and subordinate-level visual classes, the contribution of view-specific representations of shape would seem to be modest indeed (Biederman, Subramaniam, Bar, Kalocsai, & Fiser, 1999).

2. Contribution of resolution difficulties and near accidents to rotation costs when distinctive GSDs were present

Hayward and Tarr (1997) and Tarr, Williams, Hayward, and Gauthier (1998) reported rotation costs for images that, in their subjective judgments, were 'extremely easy to discriminate'. Subjective judgments of ease of recognition can dramatically underestimate differences in perceptibility (Biederman, Hilton, & Hummel, 1991). However, near accidents and resolution variations could have contributed to the modest rotation costs they observed in their studies, as described below. Moreover, their stimuli were hardly on an extreme of discrimination ease. Compare, for example, Hayward and Tarr's (1997) two-part novel objects with a set of stimuli that could be composed of the two-part novel original objects in Biederman and Bar (1999) (The same-different matching task in Biederman and Bar's experiment required detection of subtle changes of one of the geons from the same original object rather than between different original objects). Different original objects from the Biederman and Bar (1999) experiment would clearly be far easier to distinguish than those of Hayward and Tarr's, with little of the difficulty in segmenting the smaller geon from the larger geon in many of the poses shown in Hayward and Tarr. If Hayward and Tarr's objects are to be regarded as 'extremely easy to discriminate', should a selection of original objects from Biederman and Bar's (1999) experiment be regarded as 'extremely, extremely easy to discriminate?' Interestingly, Hayward and Tarr do not take issue with Biederman and Bar's (1999) expectation that little or no rotation costs would be observed if subjects had to discriminate among the original objects.

Tarr et al. (1998) reported rotation costs for the single geon objects that were identical to those of Biederman and Gerhardstein (1993, Experiment 3). Biederman and Gerhardstein (1993) did not take pains to equate all their stimulus pairs in similarity at all orientations — indeed, they remarked on the presence of near accidents in their stimuli. Biederman and Gerhardstein (1993, 1995) also explicitly noted the possibility that resolution variations and near accidents could have contributed to the small rotation effects that they observed. The marked variation in false alarm rates, from 0 to 100%, in the confusion matrix in Biederman and Gerhardstein's (1993) Exp. 3 is objective evidence that there was considerable variability in the similarity from stimulus pair to stimulus pair. It would be instructive to know whether in the data of Tarr et al. (1998) the contribution to the rotation costs (RTs in their case) did *not* come from those stimulus pairs where resolution was difficult.

Concerning the high error rates in the Biederman and Gerhardstein's (1993) Exp. 3, Hayward and Tarr write: 'Since we presume the goal of any theory of recognition is *accurate* recognition, it seems more likely that the poor performance obtained by Biederman and Gerhardstein does not reflect standard recognition processes'.

We disagree. The goal of a theory of biological recognition is recognition *behavior*. People make errors, particularly when attempting to discriminate small differences under brief, masked exposures. Geon Theory's assignment of a component of task difficulty to failures of resolution is entirely consistent with a vast body of psychophysical literature. If the discrimination of a contour, say, was required to distinguish between two objects, then difficulty in discriminating that contour should be reflected in object recognition performance.

3. Many of the studies showing rotation costs used rotation angles where resolution changes could have contributed to rotation costs

4. Reflections and 180° rotations that restore the object's parts show relatively smaller costs, in contrasts to expectations of view-based theories

Biederman and Bar (1999) did not maintain, as argued by Hayward and Tarr, that larger rotations, in general, were more appropriate for the testing of rotation effects. We merely noted that 180° rotation (or reflection) of bilaterally symmetrical objects offer a

clear test of rotation costs where resolution variations are not at issue. (Hayward and Tarr are correct in pointing out that many orientations, e.g. the front and back of an airplane, would provide different part structures from a 180° rotation. We should have specified views that restore the parts, as side views of such objects.) The invariance of performance for such reflections — as reported by Biederman and Cooper (1991a,b) — is a clear challenge to the metric templates espoused by a number of theorists. As noted by Biederman and Gerhardstein (1993) cost functions over rotation angle are often nonmonotonic, declining from 90 to 180° for near side views of bilaterally symmetrical objects. Thus *smaller* costs are obtained than would be predicted by rotations of 30° to 90° where occlusion and foreshortening do affect the images. A rotation of 30° would, on average, produce smaller degrees of self-occlusion and 'near accidents', than one of 90°, but these resolution effects are reduced as rotation of bilaterally symmetrical objects increases towards 180°. So the cost functions follow the resolution difficulties — not the rotation angle. These resolution and near accident effects are difficult to avoid and, insofar as they depend on the initial arbitrarily defined 0° view, there is no simple relation between rotation angle and these effects. It was for precisely this reason that Biederman and Bar abandoned the use of a constant rotation angle for all objects and employed different rotation angles, ranging from 20 to 120°, for each of their objects. Hayward and Tarr assert that in their (1997) study "… there *were no changes in the visible parts* of any of the objects shown in the study." (Italics theirs.) This, of course, is impossible, as there will be some foreshortening and lengthening of surfaces and contours with depth rotation.

Hayward and Tarr argue that the Tarr et al. (1998) study showed 'costs across precisely the range of viewpoints advocated by Biederman and Bar'. But this is not true. We advocated viewpoints that were a 180° rotation from the 0° viewpoint. Tarr et al. (1998) used viewpoints that 'spanned 180°; 90° in each direction from the training viewpoint'. (Hayward and Tarr, point 4). So their maximum viewpoint change was not 180°, but 90° — the view change that, on average, tends to maximize occlusion and accidental effects relative to the 0° view.

Hayward and Tarr note that some very recent models (e.g. Riesenhuber & Poggio, 1999, but they could have cited an earlier model by Mel, 1997) are not of the metric template kind that Biederman and Bar address. This is true. In positing a central role to view-invariant features and discarding a 2D coordinate space, such models come closer to what we have been advocating all along. One wonders about the findings that motivated this drastic change. Hayward and Tarr state that the older variety of template models, e.g. Poggio and Edelman, 1990, would predict reflection invariance but this was not a characteristic of that model and certainly not a characteristic of mental rotation.

5. Biederman and Bar (1999) [and Biederman and Gerhardstein (1993, 1995)] noted that "… rotation in depth tends to produce drastic changes in the 2D image that can differentially affect the perceptibility of the parts" (Biederman & Bar, p. 2896)

Hayward and Tarr erroneously interpret this statement to mean that rotations, in general, make perceptibility more difficult. The only reasonable interpretation of our statement, especially given the context and that 0° views were balanced, is that if some parts were more readily resolved at one view and other parts at other views, then rotation costs might be produced by these differences in perceptibility at the different orientations, as noted in our response to the previous point.

In this section, Hayward and Tarr assert that similar orientation differences produced similar orientation costs for different objects. We here simply repeat from Biederman and Bar (1999), the findings of Biederman and Bar (1998) in matching bent paper clip stimuli: There were gigantic differences in the rotation costs and false alarm rates from stimulus pair to stimulus pair at the same orientation differences. These differences were readily interpretable in terms of the nonaccidental feature differences (produced by accidental configurations of the clips, as discussed by Biederman & Gerhardstein, 1993). Nonaccidental differences between two views of the same clip produced high miss rates. When such differences were absent in views of two different clips, false alarm rates were extremely high (up to 85%).

6. Rendered images. "This statement clearly indicates that Biederman and Bar believe that rendered images are somehow less appropriate stimuli than line drawings for studying human visual recognition"

Biederman and Bar's statement was that "rendered images, as compared to line drawings, typically have lower contrast and illumination and shadow contours that can increase the difficulty of determining the orientation and depth discontinuities important for resolving the geons" (p. 2896). To the extent to which this shape information may be difficult to resolve, other sources of information may be employed, as discussed previously. Our statement is absolutely clear. It was motivated by our observations that a distinguishing part in rendered images of novel objects was sometimes difficult to resolve at brief presentations. This led to our use of somewhat longer exposure durations relative to previous studies insofar as we were using rendered images. With the rendered novel images (Biederman & Bar, 1998), the longer exposure durations, in fact, led to reduced rotation costs.

Nothing in Biederman and Bar's statements indicated that we think that line drawings have more ecological validity than rendered images (and those, of course, are

not identical to real objects). Hayward and Tarr go on to question whether a line drawing marking the orientation and depth discontinuities can be extracted by the human visual system and cite an experiment by Sanocki, Bowyer, Heath, and Sarkar (1998) as evidence for their point. But the Sanocki et al. experiment merely showed that subjects are not as able to recognize the output of a particular edge detector, that devised by Canny (1986), compared with original photography. The Canny edge finder decidedly does *not* produce a line drawing of the depth and orientation discontinuities of an object but, instead, misses some of the desired discontinuities and produces additional edges from luminance and texture changes. Contrary to Hayward and Tarr's claims, people have no trouble pointing to the orientation and depth discontinuities of novel objects. Well-designed assembly instructions for toys and other products almost always include line drawings of the parts, despite the additional expense incurred in making such drawings. Moreover, a monkey IT cell responding to a color photograph of an object, generally maintains its tuning preference to line drawings of that object (Kovács, Sáry, Köteles, Chadaide, & Benedek, 1998) and recent work has indicated that monkeys trained to respond differentially to a set of ten color photographs of objects, readily transfer to line drawings of those objects (G. Kovács, personnel communication, 2000). Hayward and Tarr also cite the Biederman and Ju (1988) study showing equivalence in identification performance in line drawings and color photography. Ignoring the inconsistency of this citation with their prior point that line drawings cannot be extracted from natural images, Biederman and Ju had to undertake several tries with a professional photographer to reduce (but not eliminate) the problem of part resolution in their color photographs.

7. Contribution of transient shifts to rotation costs

Biederman and Bar (1999) noted that in sequential same-different matching of depth-rotated objects, any difference in the local areas occupied by the first and second images can result in a signal that something has changed. Only on 0°-SAME trials is there no change in the spatial positions. The absence of any S1–S2 change could thus serve as a completely reliable cue to respond SAME, which, by artifactually lowering RTs at 0°, would contribute to a positive slope over rotation angle. Biederman and Bar (1998) found precisely such an effect. When they translated S2 (with respect to S1) on all trials, RTs for SAME responses at 0° increased compared to an untranslated condition, producing a reduction in slopes (costs) over rotation angle, as there was little effect of shifting on the rotated trials. Hayward and Tarr report that Hayward and Williams (2000) found that with a translation there was still a monotonically increasing rotation cost. As Hayward and Williams did not use a no-shift control, this description of their results omits the critical point of whether the slope was reduced by the translation.

Biederman and Bar (1999) noted that, in contrast to Tarr's (1995) theoretical position of template extrapolation-interpolation (at small rotation angles)/mental rotation (at larger rotation angles), the rotation cost function is positively accelerated, not negatively accelerated (Tarr, 1995). That is, the difference between 0 and 30° produced greater costs than between 30 and 60°. The relatively steeper slope near 0° could have been produced by a transient artifact. In asserting that the rotation costs were monotonic, Hayward and Tarr fail to address our point that template transformations would lead to positive, not the obtained negative, acceleration.

In considering a possible neural basis of the transient shift signal, Biederman and Bar cited a result of Nowak and Bullier (1997) who reported a fast magnocellular transient response in IT cells. Biederman and Bar (1999) (p. 2887) explicitly noted that: 'Because of the intervening mask, the transient in the present case would have to be one which was a function of the difference between S2 and an actively maintained representation of S1. Active maintenance would be necessary to avoid the disruption of the mask. Indeed, this is the subjective impression of what one is doing when performing the task'. Hayward and Tarr ignore our qualification and repeat our point that the mask would furnish a transient. It is an empirical issue, of course, as to whether active maintenance of S1 in a sequential matching task: (a) reduces the transient from the mask; and/or (b) enhances a transient from any S1–S2 differences. There is, indeed, no evidence for the neural basis of this hypothesis: Not because investigators looked for such an effect and failed to find it, but rather because no one, as yet, has looked.

Acknowledgements

Supported by ARO DAAH04-94-G-0065 (to IB), and the McDonnell-Pew program in Cognitive Neuroscience, 99-6 CNS-QUA.05 (to MB). E-mails: bieder@usc.edu and bar@nmr.mgh.harvard.edu.

References

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115–147.

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research*, *39*, 2885–2899.

Biederman, I., & Bar, M. (1998). Same-different matching of depth-rotated objects. *Investigative Ophthalmology & Visual Science*, *39*, 1113 (Abstract, ARVO).

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., & 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., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology*: *Human Perception and Performance*, *21*, 1506–1514.

Biederman, I., & Ju, G. (1988). Surface vs. edge-based determinants of visual recognition. *Cognitive Psychology*, *20*, 38–64.

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., Hilton, H. J., & Hummel, J. E. (1991). Pattern goodness and pattern recognition. In J. R. Pomerantz, & G. R. Lockhead, The perception of structure (Chapter 5, pp. 73–95). Washington, D.C.: APA.

Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P, & Fiser, J. (1999). Subordinate-Level object classification reexamined. *Psychological Research*, *62*, 131–153.

Canny, J. F. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8*, 769–798.

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.

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.

Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science*.

Kovács, G., Sáry, G., Köteles, K., Chadaide, Z., & Benedek, G. (1998). Effect of surface attributes on the shape selectivity of inferior temporal neurons. Poster presented at the Meetings of the Society for Neuroscience, Los Angeles, CA, November.

Mel, B. W. (1997). SEEMORE: combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition. *Neural Computation*, *9*, 777–804.

Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In J. Kaas, K. Rockland, & A. Peters, *Extrastriate cortex*, *cerebral cortex*, *vol. 12*.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, *343*, 263–266.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.

Sanocki, T., Bowyer, K. W., Heath, M. D., & Sarkar, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology*: *Human Perception and Performance*, *24*, 340–349.

Tarr, M. J. (1995). Rotating objects to recognize them: a case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, *2*, 55–82.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, *1*, 275–277.