a Inc. • http://neurosci.nature.com

Three-dimensional object recognition is viewpoint dependent

Michael J. Tarr[1], Pepper Williams[2], William G. Hayward[3] and Isabel Gauthier[4]

[1] *Brown University, Department of Cognitive and Linguistic Sciences, Box 1978, Providence, Rhode Island 02912, USA*

[2] *University of Massachusetts, Department of Psychology, 100 Morrissey Boulevard, Boston, Massachusetts 02125-3393, USA*

[3] *University of Wollongong, Department of Psychology, Northfields Avenue, Wollongong, NSW 2552, Australia*

[4] *Yale University, Department of Psychology, PO Box 208205, New Haven, Connecticut 06520-8205, USA*

Correspondence should be addressed to M.J.T. (Michael_Tarr@brown.edu)

The human visual system is faced with the computationally difficult problem of achieving object constancy: identifying three-dimensional (3D) objects via two-dimensional (2D) retinal images that may be altered when the same object is seen from different viewpoints[1]. A widely accepted class of theories holds that we first reconstruct a description of the object's 3D structure from the retinal image, then match this representation to a remembered structural description. If the same structural description is reconstructed from every possible view of an object, object constancy will be obtained. For example, in Biederman's[2] oft-cited recognition-by-components (RBC) theory, structural descriptions are composed of sets of simple 3D volumes called geons (Fig. 1), along with the spatial relations in which the geons are placed. Thus a mug is represented in RBC as a noodle attached to the side of a cylinder, and a suitcase as a noodle attached to the top of a brick. The attraction of geons is that, unlike more complex objects, they possess a small set of defining properties that appear in their 2D projections when viewed from almost any position (e.g., all three views of the brick in Fig. 1 include a straight main axis, parallel edges, and a straight cross section). According to the RBC theory, a complex object can therefore be recognized from its constituent geons, which can themselves be recognized from any viewpoint.

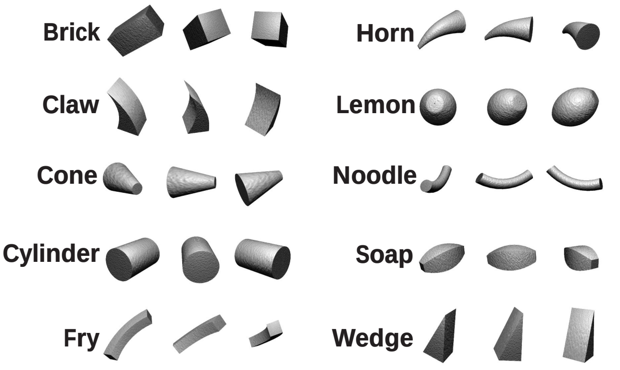

Fig. 1. Shaded images of the three views of the ten geons used in the experiments, along with names assigned in experiment 3. The leftmost figure in each row was arbitrarily designated the 0° view; the other two figures represent 45° and 90° rotations of the objects in the depth plane.

© 1998 Nature America Inc. • http://neurosci.nature.com

A fundamental assumption of RBC is that recognition of individual geons (and therefore objects composed of geons) should be equally accurate and fast when seen from almost any viewpoint (barring 'accidental' viewpoints[3]). Do humans actually recognize geons in such a viewpoint-invariant manner? Recently published reports[3,4], including single experiments using single geons, have produced inconsistent results. Our set of experiments was designed as a definitive test of RBC's postulate that geon recognition should be viewpoint invariant, including nine experiments utilizing three different tasks (sequential matching, match-to-sample and naming), two different versions of geons (line drawings and shaded images) and several other factors that might be expected to influence recognition performance.

Experiments 1a–e utilized a sequential matching task, in which two images were presented back to back and participants decided whether they depicted the same or different geons. (Trials in which different geons were presented were not of theoretical interest, so only the results of 'same' trials are discussed.) Image pairs represented geons viewed from vantage points differing by 0°, 45° or 90°. The null hypothesis, that geons were recognized without a cost for changes in

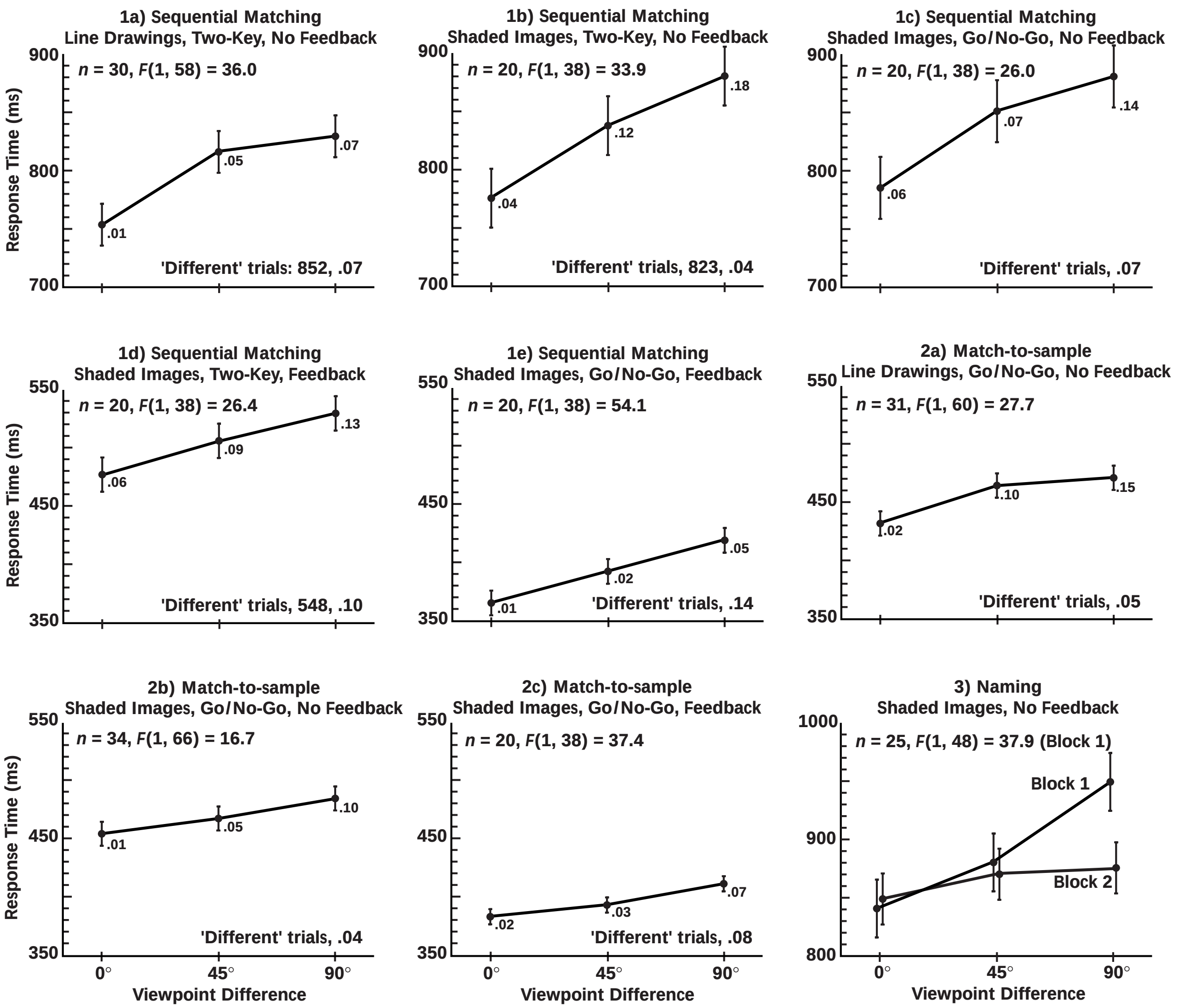

Fig. 2. Results of the psychophysical experiments. **(1a–e)** Mean response times to judge that two sequentially presented images represented the same geon; **(2a–c)** Mean response times to judge that target geon images matched a previously presented sample geon image; **(3)** Mean response times to name geon images. (See **Fig. 1** for names given to geons.) Error bars show within-participants 95% confidence intervals[13], and error rates are given beside each data point. (Error rates were not recorded in experiment 3.) Major procedural differences between experiments are given above the graphs. Experiments 1a and 2a used line drawings scanned in from ref. 3, whereas other experiments used the shaded geons shown in Fig. 1. In experiments 1a, b and d, participants pressed one key to respond 'same' and an alternate key to respond 'different,' whereas experiments 1c and e and 2a, b and c used a go/no-go procedure, in which participants pressed a key to respond 'same' or did nothing otherwise. (Participants in experiment 3 spoke geon names into a microphone.) In experiments 1d and e and 2c, participants received feedback on each trial about the accuracy and speed of their response. Numbers of participants and results of linear contrast tests are given in the upper left of each graph (all contrasts were significant at the $p < .0001$ level), and response times (in milliseconds) and error rates for trials in which the correct response was 'different' are given in the bottom right. (Reponse times for 'different' trials were not recorded in go/no-go experiments.)

© 1998 Nature America Inc. • http://neurosci.nature.com

© 1998 Nature America Inc. • http://neurosci.nature.com

viewpoint, is unsupportable for any of these five experiments. (See Fig. 2 for results of all nine experiments.)

The same conclusion holds for experiments 2a–c, which used a match-to-sample task. In this task, trials were run in blocks, where a participant saw a target geon in the 0° view, followed by three trials each of the 0°, 45° and 90° views of the target geon, interspersed with nine other geons. Again, clear viewpoint-dependent effects were found. The apparent reduction in the size of the effects here compared to experiments 1a–e was at least in part a result of practice; for the first trials in each block, the difference between 0° and 90° views averaged 52 ms (averaged over all 85 participants in experiments 2a–c), whereas for the third trials with each view, the difference averaged only 22 ms. This interaction of trial number and viewpoint is reminiscent of results from other experiments[5], in which viewpoint effects were stronger in initial than in later blocks of trials. Indeed, such practice effects may help explain why our findings are at odds with those of experiment 4 of Biederman and Gerhardstein[3], in which the procedure was similar to our match-to-sample experiments but no viewpoint effects were found.

The same phenomenon was observed for the naming task used in experiment 3, in which participants first learned labels for the 0° view of each geon, then were asked to name 0°, 45° and 90° views in two subsequent blocks. The effect of viewpoint difference was highly significant in the first block but greatly reduced in the second block, as the subjects learned the new viewpoints.

Although viewpoint effects in each experiment were significant, it is conceivable that these overall patterns were the result of a small subset of anomalous participants and/or stimulus items. We tested for this contingency with nonparametric sign tests, which indicated that across the nine experiments, 86% of participants and an average of 9.3 of the 10 geons were faster for 0° viewpoint changes than 90° viewpoint changes, 76% of participants and 8.2 geons were faster for 0° viewpoint changes than for 45° viewpoint changes and 70% of participants and 7.3 geons were faster for 45° viewpoint changes than for 90° viewpoint changes (all $z > 4.43$, all $p < .001$).

The viewpoint-dependent effects revealed in these experiments could not have been due to certain views being inherently easier or harder to process. In the sequential matching procedure of experiments 1a–e, all pairwise combinations of views were tested. That is, the 0° viewpoint difference condition includes trials testing all three of the 0°, 45°, and 90° views in Fig. 1, the 45° viewpoint difference condition includes trials testing 0°–45° view combinations and 45°-90° view combinations in both possible orders and the 90° viewpoint difference condition includes 0°–90° trials and 90°–0° trials. Thus, the decrease in performance from 0° to 45° to 90° conditions results from the changes of viewpoint in the latter two conditions, not from the particular views that were tested.

Like most theories of object recognition, RBC[2] attempts to explain how the visual system achieves relatively consistent identification of 3D objects even though their 2D projections vary widely when seen from different viewpoints. RBC's *modus operandi* is to decompose objects into collections of geons and to propose that geons are equally recognizable from almost any viewpoint. If this proposition were true, then experimental participants' ability to decide that two images represent an identical geon should remain constant when the two images are taken from different viewpoints. This hypothesis was decisively rejected nine times in the experiments presented here. It is still theoretically possible that objects are recognized by first parsing them into sets of geons (or some other type of primitive). However, our experiments demonstrate that structural descriptions based on geons cannot be recovered from images in a viewpoint-invariant manner, as the recognition of geons themselves is viewpoint dependent.

More generally, our findings of substantial and robust viewpoint effects for extremely simple 3D volumes argue against any scheme proposing viewpoint-invariant representations within the brain as a basis for object recognition[1]. This conclusion is consistent not only with recently reported behavioral experiments using multi-part objects[4,6–9], but also with neurophysiological studies. For example, neurons responsive to human faces have been found in the macaque superior temporal sulcus, with the majority of these cells preferential for specific views of faces[10]. Neurons in inferior temporal cortex show the same viewpoint specificity for novel 3D objects that monkeys were trained to recognize[11]. Taken together, these neurophysiological and behavioral findings offer persuasive evidence that object recognition is a viewpoint-dependent process. This perspective is embodied in theories assuming that collections of features, surfaces, parts or entire images of objects are encoded in a viewpoint-specific manner[12] and that object recognition processes are based on the similarity between encoded and perceived images. Objects seen from viewpoints increasingly different from learned views will tend to project increasingly less-similar images, so view-based theories provide a natural account for the types of viewpoint effects found here.

Note: Further details of methods may be found on the Nature Neuroscience web site at http://neurosci.nature.com/web_specials/

Acknowledgements

This research was supported by an Air Force Office of Scientific Research Grant. We thank Jay Servidea and Jaymz Rosoff for their assistance in running the psychophysical studies.

RECEIVED 4 MAY: ACCEPTED 4 JUNE 1998

1. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, 1982).
2. Biederman, I. *Psychol. Rev.* 94, 115–147 (1987).
3. Biederman, I. & Gerhardstein, P. C. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 1162–1182 (1993).
4. Hayward, W. G. & Tarr, M. J. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 1511–1521 (1997).
5. Jolicoeur, P. *Mem. Cognit.* 13, 289–303 (1985).
6. Bülthoff, H. H. & Edelman, S. *Proc. Natl. Acad. Sci. USA* 89, 60–64 (1992).
7. Humphrey, G. K. & Khan, S. C. *Can. J. Psychol.* 46, 170–190 (1992).
8. Tarr, M. J. *Psychonomic Bull. Rev.* 2, 55–82 (1995).
9. Tarr, M. J., Bülthoff, H. H., Zabinski, M. & Blanz, V. *Psychol. Sci.* 8, 282–289 (1997).
10. Perrett, D. I. *et al. Proc. R. Soc. Lond. B* 223, 293–317 (1985).
11. Logothetis, N. K., Pauls, J. & Poggio, T. *Curr. Biol.* 5, 552–563 (1995).
12. Poggio, T. & Edelman, S. *Nature* 343, 263–266 (1990).
13. Loftus, G. R. & Masson, M. E. J. *Psychonomic Bull. Rev.* 1, 476–490 (1994).

© 1998 Nature America Inc. • http://neurosci.nature.com