VISUAL COGNITION, 2005, 12 (5), 705–719

Psychology Press
Taylor & Francis Group

Combining disparate views of objects: Viewpoint costs are reduced by stereopsis

Darren Burke
Macquarie University, Sydney, Australia

An issue of central concern in the object recognition literature is whether changes in the viewpoint from which an object is depicted produces systematic costs in performance, or whether performance is (largely) unaffected by such changes. This issue has generated a vigorous and lengthy debate because viewpoint-dependent or viewpoint-independent performance has been seen as a reflection of the underlying object representations. The current experiment shows that the effect of viewpoint differences between objects is strongly affected by whether or not they are depicted with stereoscopic depth, a result that is predicted by neither of the main approaches to object recognition. Instead, it is proposed that viewpoint costs in object recognition experiments are a function of the extent to which the information a subject is provided with generalizes across views, without this holding any necessary implications for the nature of the underlying object representations.

The ease and speed with which we recognize objects in our everyday lives belies the complexity and sophistication of the feat we are achieving. One serious problem to be overcome is the fact that the retinal image produced by most objects varies dramatically with viewpoint (consider a cat viewed from the side or from directly in front). How we deal with these changes has been a major focus of object recognition research, with two main approaches emerging. The most obvious way to account for viewpoint generalization is to propose that a single view of an object is (usually) sufficient to extract its three dimensional structure, and that the object is represented in an allocentric, viewpoint-independent, framework, so that new views will simply activate the same representation (Biederman, 1987; Marr & Nishihara, 1978). A less computationally intensive approach is to assume that a single view of an object produces only a view-specific, egocentric, representation, and that generalization to new views is based on some kind of extrapolation or interpolation (Poggio & Edelman 1990;

Please address all correspondence to: Darren Burke, Department of Psychology, Macquarie University, Sydney, NSW, 2109, Australia. Email: dburke@vision.psy.mq.edu.au.

I would like to thank William Hayward, Brian Stankiewicz, and two anonymous reviewers for comments that were helpful in the production of this paper.

http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280444000463

Tarr & Pinker, 1989). According to allocentric approaches, a new view activates the same representation as an old view (provided the same structural information is available in the two views), so changes in viewpoint should incur no recognition costs (in terms of reaction time or errors). Egocentric approaches predict viewpoint costs proportional to the difference in the views, as a consequence of the extrapolation or interpolation required.

Because viewpoint dependent (VD) or viewpoint independent (VI) performance is thought to reflect the fundamental nature of the underlying object representations, a lively and protracted debate has centred on which kind of performance is characteristic of human object recognition (for example; Biederman & Bar, 1999, 2000; Biederman & Gerhardstein, 1993, 1995; Bulthoff & Edelman, 1992; Hayward & Tarr, 1997, 2000; Tarr & Bulthoff, 1995, 1998). Although there have been attempts to reconcile these views by suggesting that object recognition may use both VD and VI information, either in separate neural pathways (Burghund & Marsolek, 2000; Vanrie, Beatse, Wagemans, Sunaert & Van Hecke, 2002) or in a simple additive fashion (Foster & Gilson, 2002), and there has even been a recent suggestion that the viewpoint debate is essentially over (Hayward, 2003), there are clear signs that consensus may not be imminent. For example, in a recent review of the role of expertise in object recognition, Tarr and Cheng (2003) claim that a single, VD object recognition system is sufficient to account for all levels of visual categorization, and although not focused exclusively on viewpoint costs, a similar debate continues on the nature of object representations (Edelman & Intrator, 2003a, 2003b; Hummel, 2003).

We may be in a better position to reconcile these conflicting approaches if we acknowledge that VD or VI *performance* on an object recognition task does not necessarily hold any direct implications for the nature of the underlying object representations (a point also made by viewpoint independent theorists; Bar, 2001; Stankiewicz, 2002, 2003). All object recognition experiments are, for reasons of control and tractability, simulations of real world object recognition, and so they necessarily provide the observer with only a subset of the information that is normally available to recognize real objects. It may be that the particular subset selected in any given experiment is a better predictor of performance than the putative underlying object representations. From this perspective, viewpoint costs are a function of the extent to which the experimental procedure in question provides the observer with *information* that is generalizable across viewpoints, and so they do not, alone, hold any implications for how the object is represented.

For example, perhaps the strongest evidence for Biederman's Geon Structural Description theory, the main VI approach to object recognition, is that viewpoint costs are much more modest when novel objects can be discriminated on the basis of parts that differ in terms of non-accidental property than when they differ only on the basis of a metric property (like the width, length or precise angle of attachment of parts) (Biederman & Bar, 1999). Since a non-accidental property is, *by definition*, a property that does not change with changes in viewpoint, this result may not be entirely surprising to those outside the viewpoint debate (although some VD theories explicitly claim that such properties are unimportant—Edelman, 1995—and no VD theories emphasize such properties). Similarly, Vanrie et al. (2002) have recently shown that when objects are discriminable on the basis of an ''invariant feature'' (a categorical difference in the angle at which each of the parts joins the main body), which is detectable no matter what angle the object is seen from, then performance is, indeed, VI. The fact that performance on a traditional mental rotation task, using stimuli that differed only in that they were mirror reflections, showed strong effects of viewpoint, and that fMRI revealed different patterns of activation for the two tasks, led them to propose that there are two routes to object recognition, one of which (in the dorsal stream) is VD and other of which (in the ventral stream) is VI. The second claim is inconsistent with results from Gauthier, Hayward, Tarr, Anderson, Skudlarski and Gore (2002), who also used fMRI to compare activity associated with mental rotation and object recognition (without an invariant feature), and found strongly VD performance in both tasks which correlated with activity in the dorsal stream for mental rotation but in the ventral stream for object recognition. Together, these studies reinforce the idea that object recognition is a ventral function (and that mental rotation is something different), but that it can proceed in a VI or VD manner, depending on the information made available to the observer to discriminate between the objects.

The strongest evidence for view-specific approaches is that the great majority of object recognition tasks and stimulus sets have produced VD performance (Lawson, 1999), which becomes less VD as observers are exposed to new views (Tarr & Pinker, 1989). According to such approaches, exposure to new views creates new view-specific representations, a claim that is consistent with the known neurophysiology (Logothetis & Sheinberg, 1996), and so any novel view is closer to a stored view, reducing viewpoint costs. While exposure to more than one view of an object does provide the opportunity to form new view-specific representations, it also provides more *information* about the object than a single view. New views obviously provide explicit information about what the object looks like from different views, and they probably help the subject to work out its three dimensional structure. The reduced viewpoint costs could be a consequence of the improvement in the quality of this information.

If, as suggested above, viewpoint costs might be reduced by providing more information about the 3D structure of an object, then there is a much more straightforward means of providing that information than training observers with multiple views—stereopsis. There have, in fact, been two previous studies that are widely reported to have failed find a reduction in viewpoint costs when stereo information was provided in an object recognition task (Edelman & Bulthoff, 1992; Humphrey & Khan, 1992), but in neither case is the evidence compelling. Because they considered reaction time (RT) to be of less theoretical importance, Edelman and Bulthoff (1992) only reported error rates, and found a general benefit for stereo conditions (participants made fewer errors overall), but no interaction between viewing condition and viewpoint cost. It is impossible to decide whether this is a consequence of a speed-accuracy tradeoff without RT data. In the analysis of their error data, Humphrey and Khan (1992) in fact *did* find a significant three-way interaction between viewing condition (binocular or monocular), training view and test view (viewpoint costs were smaller in the binocular condition), but decided that this was probably the consequence of a speed-accuracy tradeoff.

The current experiment used digital images of real, haphazardly bent paperclips (see Figure 1), to re-examine the possible role of stereoscopic information in helping observers to generalize between views. Participants were

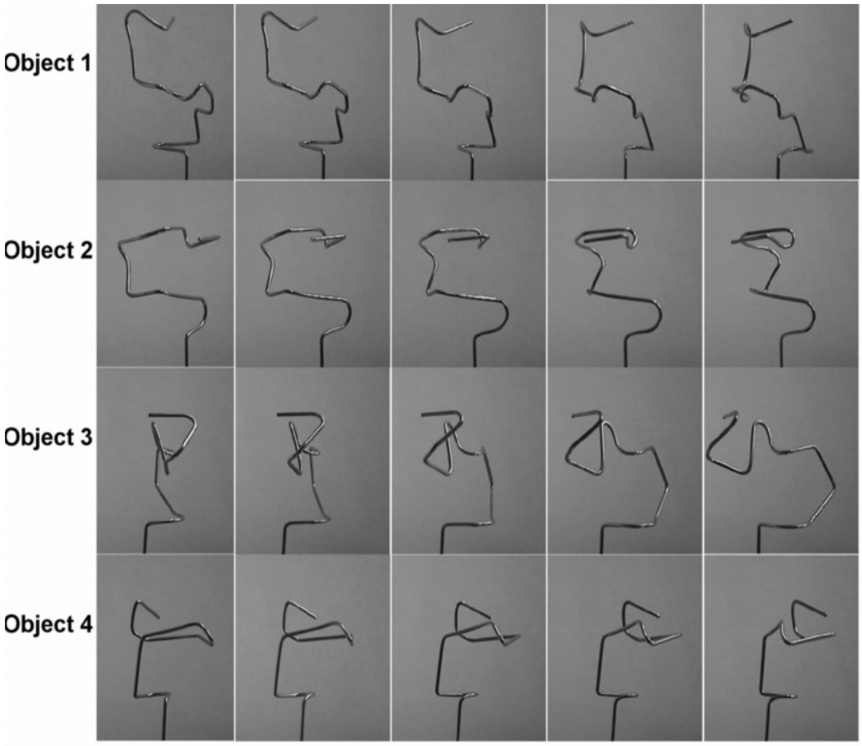

Figure 1. The left eye views of each of the objects at each of the viewpoints used in the experiment.

exposed to two objects in succession (separated by a mask) and were asked to decide whether the second object was the same as the first, or was a different object. In the main experiment, all objects were presented as pairs of images in a prism stereoscope. Half of these presentations showed slightly different views in each eye, and so stereopsis provided the full 3D structure of the object to subjects. The other half of the presentations contained the same view of the object in each eye (synoptic presentation). This mode of presentation contains *no* stereoscopic information about the structure of the object (and so simulates a distant object). The usual mode of presentation in object recognition experiments, binocularly viewing a single object on a computer monitor, in fact contains explicit stereoscopic information that the object is *flat*, which conflicts with the other (monocular) 3D information. Because a stereoscopic view of an object necessarily contains two different *views* of the object (albeit rather similar ones), there is a possibility that this is sufficient to produce a reduction in viewpoint costs (Tarr & Pinker, 1989), rather than stereoscopic information *per se* being responsible for any observed differences. For this reason a control experiment was also run, which was identical to the main experiment except that subjects viewed the stimuli without the stereoscope, and so saw two horizontally separated views of each object (exactly as in Figure 2).

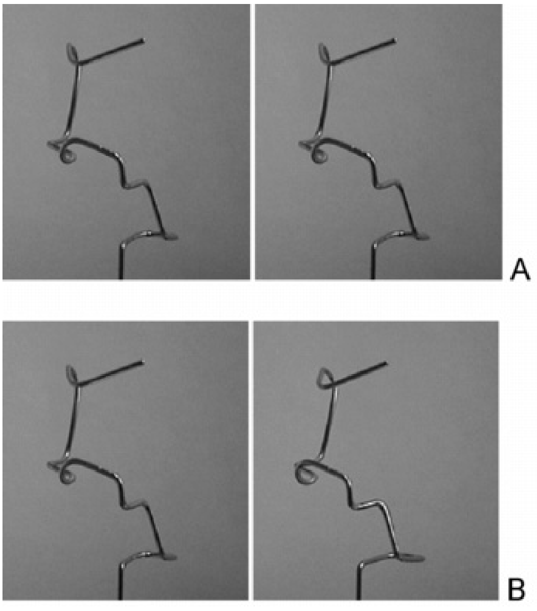

Figure 2. Stereopairs of object 1 at the 80° viewpoint. A is the Synoptic view and B is the Stereo view. The stereopairs can be fused by looking through the page.

To anticipate the results, participants were much better able to generalize between views when the objects were Stereo than when they were Synoptic (they were faster and made fewer errors), but only with the benefit of the stereoscope, suggesting that stereoscopic information is necessary to produce the difference. This finding is not predicted by either of the main approaches to viewpoint generalization outlined earlier, but is perfectly consistent with the idea that viewpoint costs are a function of the information available to perform the task.

METHODS

Participants

Eight students and two staff members (including the author) of Macquarie University participated in the main experiment. Three were female.

Eight students of Macquarie University participated in the control experiment. Four were female.

Materials and procedure

Four paperclips were bent into different shapes and mounted on a rotatable platform to be photographed. The camera was positioned 13.5cm from the centre of the platform. Each object was photographed twice (from views 3.5 cm apart, to produce stereo-pairs), at 5 viewpoints, separated by 20° (see Figure 1). In the main experiment a prism stereoscope was mounted 9cm from the monitor, so the stereo-halves subtended $28° \times 33°$ of visual angle. In the experiment subjects were presented with both the left-eye and right-eye view of the object, so that they saw it in Stereo, or with the left-eye view presented to both eyes, producing a Synoptic depiction of the object (see Figure 2). In the control experiment the stereoscope was not used, so participants saw two adjacent views of each object on each presentation. Stimuli were presented on an NEC RGB monitor, driven by a Macintosh G3. Stimulus presentation and data collection were controlled by RSVP (Williams & Tarr, no date).

Participants performed a successive, same-different task, in which they saw the first object for 2500 ms, which was replaced by a pattern mask (made up of randomly shuffled segments of the object images) for 500 ms, and then the second object was shown until a response occurred or 5000 ms elapsed. They were instructed to respond as quickly and as accurately as possible, and to press the ''same'' key even if the second object was shown from a different viewpoint, and to respond ''different'' only if the two presentations were of different objects. Pictorial examples of what was meant by this were provided.

On ''same'' trials, each viewpoint of each object appeared as the first object and as the second object, in every possible combination (0° as the first object, paired once with itself and once with each of the other viewpoints, etc.). There are 25 such combinations, 4 different objects, and each of these was shown with both objects in the sequence depicted Stereo or Synoptic, producing a total of 200 same trials. Two hundred ''different'' trials were produced by using essentially the same trial structure, but replacing the second object in the sequence with a different object (shown from the same viewpoint as the object it replaced). Replacement objects were selected by cycling through the other three objects. This procedure ensured that each object (at each viewpoint) appeared as often as the first and the second object, and that there were equal numbers of same and different trials. It did not ensure that every possible ''different'' combination occurred (these were not analysed in any case), or that viewpoint differences were equally common (there were four times as many 40° differences as 80° differences, for example). The 400 trails were presented in a different random order to each subject. Subjects were presented with auditory feedback when they made an error.

RESULTS

Main experiment—with stereoscope

Reaction time data

Reaction time data from correct ''same'' trials were subjected to a 2 (Viewing condition; Synoptic or Stereo) × 5 (Viewpoint difference) within subjects factorial ANOVA. Data more than 2 standard deviations above the mean for any given condition were excluded from the analysis. As is suggested by Figure 3, this analysis revealed significant main effects of Viewing condition, $F(1, 9) = 15.66$, $MSE = 23886.23$ $p = .003$, and Viewpoint difference, $F(4, 36) = 14.23$, $MSE = 16757.17$, $p < .001$, and a significant interaction between these variables, $F(4, 36) = 2.82$, $MSE = 9620.27$, $p = .039$. A regression line was fit to the viewpoint difference data for each subject, and the slopes of those lines were analysed in a within subjects one-way ANOVA comparing slopes in the Synoptic and Stereo Viewing conditions. This analysis was significant, $F(1, 9) = 6.145$, $MSE = 4.145$, $p = .035$, suggesting, as plotted in Figure 4, that Stereo viewing produced shallower viewpoint costs than Synoptic viewing.

Error data

The % errors on ''same'' trials in each condition were subjected to a 2 (Viewing condition; Synoptic or Stereo) × 5 (Viewpoint difference) within subjects factorial ANOVA. As is suggested by Figure 5, this analysis also revealed significant main effects of Viewing condition, $F(1, 9) = 42.96$, $MSE = 114.74$, $p < .001$, and Viewpoint difference, $F(4, 36) = 29.28$, $MSE = 125.12$, $p < .001$, and a significant interaction between these variables, $F(4, 36) = 5.11$, $MSE = 119.08$, $p = .002$. The same type of slope analysis conducted on the RT data was also performed for the % error data, and it also revealed significantly

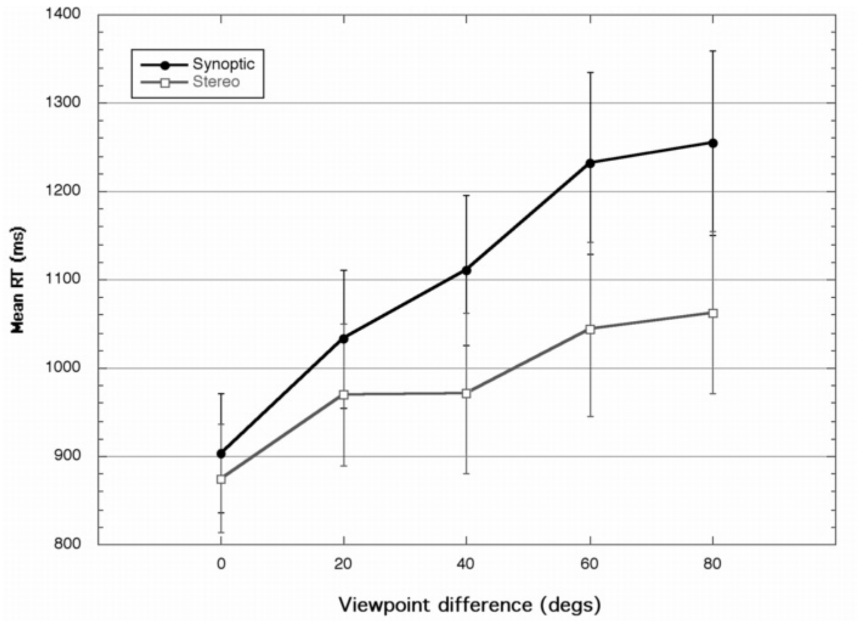

Figure 3. Mean reaction time as a function of viewpoint difference and viewing condition in the main experiment—when objects were viewed through a stereoscope. Error bars in this figure and all others represent ± 1 standard error.

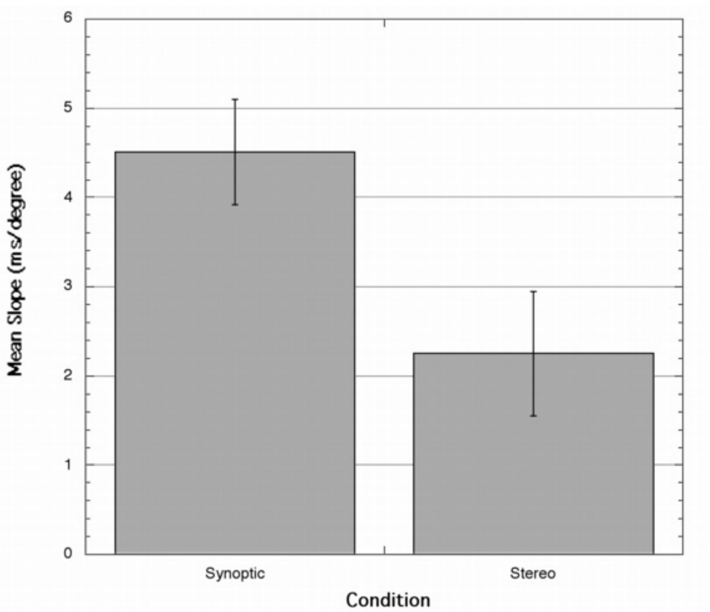

Figure 4. The mean reaction time/viewpoint difference slopes in the Synoptic and Stereo viewing conditions in the main experiment—when objects were viewed through a stereoscope.

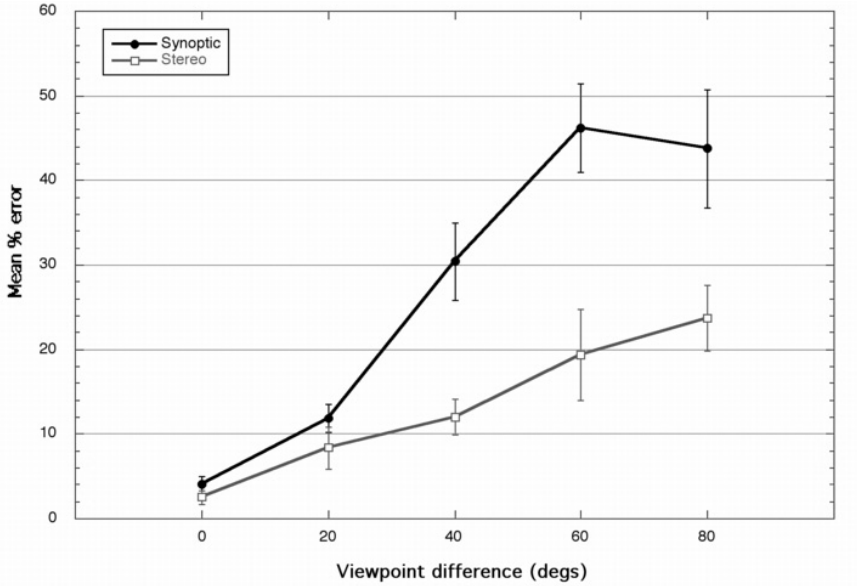

Figure 5. Mean percent error as a function of viewpoint difference and viewing condition in the main experiment—when objects were viewed through a stereoscope.

shallower viewpoint cost in the Stereo condition, $F(1, 9) = 10.43$, $MSE = 0.004$, $p = .010$. These data are plotted in Figure 6.

Control experiment—without stereoscope

Reaction time data

Reaction time data from correct ‘‘same’’ trials were subjected to a 2 (Viewing condition; Synoptic or Stereo) × 5 (Viewpoint difference) within subjects factorial ANOVA. Data more than 2 standard deviations above the mean for any given condition were excluded from the analysis. As is suggested by Figure 7, this analysis revealed no significant main effect of Viewing condition, $F(1, 7) = 1.72$, $MSE = 3798.68$, $p = .226$, a significant main effect of Viewpoint difference, $F(4, 28) = 17.54$, $MSE = 21603.97$, $p = .001$, and no significant interaction between these variables, $F(4, 28) = 0.313$, $MSE = 8850.54$, $p = .8671$. Since the main ANOVA found no differences between the viewing conditions, regression analysis was not conducted on these data.

Error data

The % errors on ‘‘same’’ trials in each condition were subjected to a 2 (Viewing condition; Synoptic or Stereo) × 5 (Viewpoint difference) within

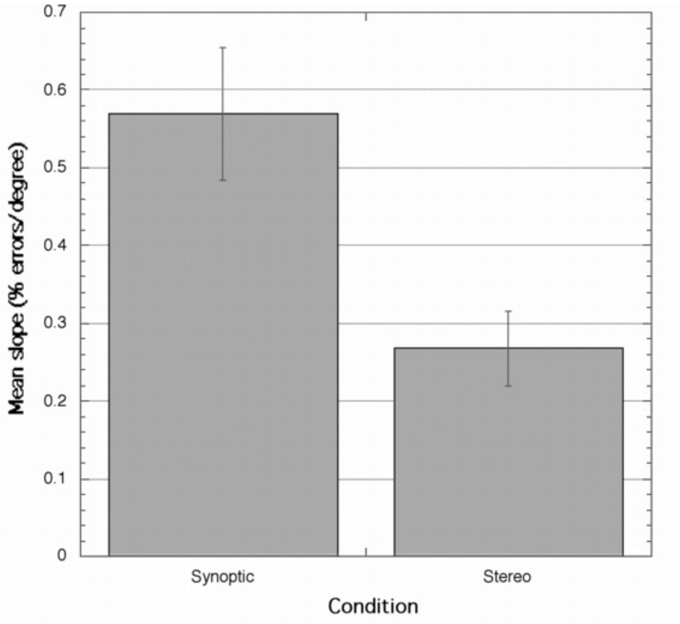

Figure 6. The mean percent error/viewpoint difference slopes in the Synoptic and Stereo viewing conditions in the main experiment—when objects were viewed through a stereoscope.

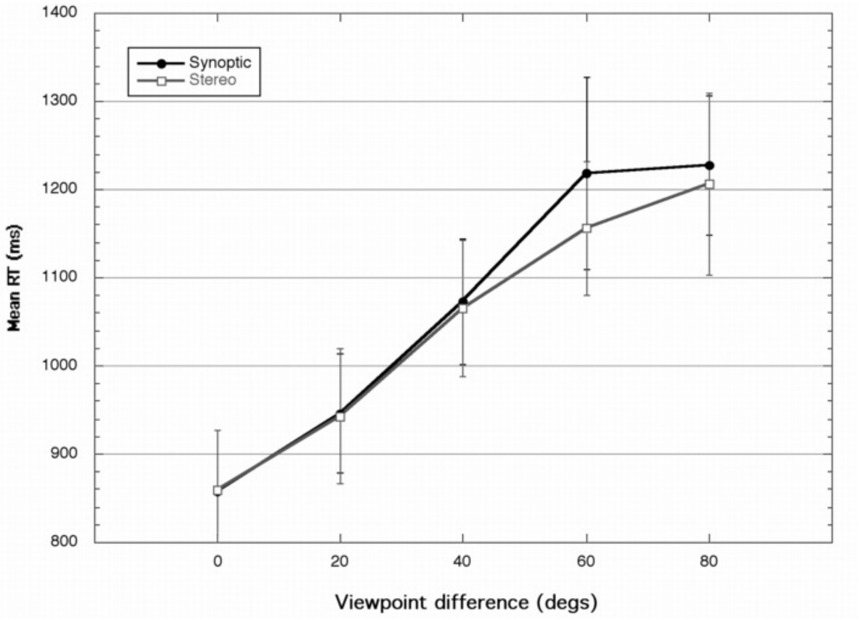

Figure 7. Mean reaction time as a function of viewpoint difference and viewing condition in the control experiment (without the stereoscope).

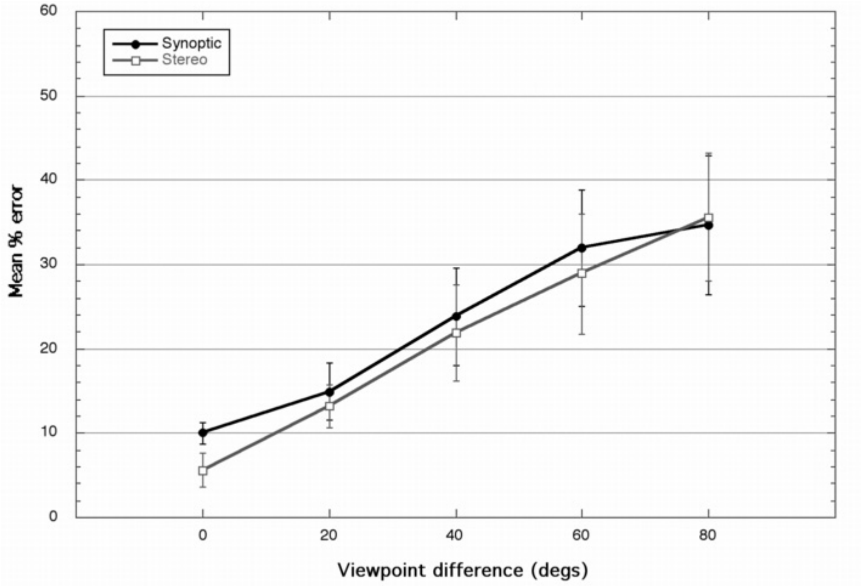

Figure 8. Mean percent error as a function of viewpoint difference and viewing condition in the control experiment (without the stereoscope).

subjects factorial ANOVA. As is suggested by Figure 8, this analysis also revealed no significant main effect of Viewing condition, $F(1, 7) = 0.608$, $MSE = 132.03$, $p < .461$, a significant main effect of Viewpoint difference, $F(4, 28) = 13.93$, $MSE = 146.30$, $p < .001$, and no significant interaction between these variables, $F(4, 28) = 0.16$, $MSE = 97.75$, $p = .957$. Again, a slope analysis was not conducted on these data.

DISCUSSION

This experiment has shown, despite previous claims to the contrary, that providing subjects with stereoscopic information about the structure of an object does help them to recognize it at different viewpoints. Viewpoint costs were approximately halved for both RT and percent error data on trials in which subjects were provided with Stereo views of the objects. This is despite the fact subjects saw the same objects depicted Stereo and Synoptic, and so might be expected to have done better in the Synoptic conditions than if they had never been given the benefit of the Stereo views. In the error data, the viewpoint cost slope in the Synoptic condition might also be an underestimate of the true costs associated with generalizing between Synoptic views, since the leveling out of performance at large viewpoint differences may be a floor effect, given that 50% is chance performance. In any case, it is clear that Stereo structural information greatly reduces (but does not eliminate) viewpoint costs, at least with the object set used in this experiment. Providing the same *views* of the objects, but without subjects being able to fuse them in a stereoscope, as in the control experiment, produced no benefit. Presumably, the two different views of each object provided in the ''Stereo'' condition in the control experiment were not different enough to help subjects generalize across viewpoints (unless they were made available to the stereoscopic system).

These data are most obviously problematic for theories in which viewpoint generalization is achieved with reference to 2D, view-specific ''snapshots'' of objects (Poggio & Edelman 1990; Tarr & Pinker, 1989). From this perspective, the current data at least mean that this process must be augmented by 3D structural information, or that the view-specific representations are, in fact, 3D. Ullman (1989) proposed a 3D, view-specific theory of object recognition, but has more recently (1998) abandoned the three-dimensionality of the representations. One problem with a 3D, viewpoint-specific theory of object recognition is that it is not really clear *how* the three-dimensionality of the representation would assist generalization between views. Presumably the view-specificity of the representations would be unaffected by their three-dimensionality, and so the same kind of normalization, interpolation or extrapolation that is characteristic of generalizing between 2D view-specific representations would be required. Such a model might, therefore, predict better performance, overall, in the Stereo than in the Synoptic viewing conditions, but it is not clear that it would predict better viewpoint generalization.

It is not entirely clear how easily VI models of object recognition (Biederman, 1987; Marr & Nishihara, 1978) could accommodate the current results. Such models appear to be consistent with the finding that viewpoint generalization is better when better 3D structural information is provided, but there are two difficulties with such a conclusion. First, despite providing explicit 3D structural information, performance was not VI. Second, although relatedly, discriminating between bent paperclips is a task to which Geon Structural Description (GSD) theory does not apply (Biederman & Bar, 1999), since it is a subordinate-level classification between objects that will generate very similar GSDs. All objects are composed of cylindrical ''geons'' attached end-to-end, and they differ primarily on the basis of metric properties—the exact angles of attachment and degrees of curvature. Even were the model applied to this situation, it would not necessarily predict better viewpoint generalization in the Stereo condition since a GSD (if one can be extracted) is already 3D. If a GSD can *only* be extracted in the Stereo case, because only then is the 3D structure of the object unambiguous, then it would presumably support VI performance.

The data from the current experiment (and those from previous studies) are consistent with the less ambitious claim that viewpoint costs are a function of the extent to which the information the subject is provided with generalizes across viewpoints. When asked to discriminate between a set of complex, novel objects, subjects presumably use whatever information is helpful, as they almost certainly do when deciding what an object is in everyday settings. Stereopsis provides them with unambiguous information about the actual 3D structure of the objects, a property that is obviously much less affected by changes in viewpoint than is the 2D view that happens to be available from any given vantage point. The claim that subjects use whatever information might help them to recognize an object is also consistent with recent data from Simons, Wang, and Roddenberry (2002), who showed that viewpoint costs were reduced if the two views of the object were produced not by simply presenting the object at a new viewpoint, but by having the subject *walk* to a new viewing position.

The fundamental difference between everyday object recognition and object recognition experiments (both perceptual and neurophysiological) is that many of the cues that differentiate objects in the real world are absent in experiments (characteristic colours, motions and sizes, for example). The perfectly sensible rationale for removing these cues is that the experiment would reduce to an investigation of colour or motion perception, and so would not provide any insights about object perception. But it is not clear that recognizing an ant, for example, based only on its size, spatial position and characteristic movements, should be excluded as an instance of object recognition because it does not rely on shape information, and there is a danger that presenting immobile subjects with static, grey-scale, isolated, 2D representations of novel objects places them in a situation that is so unlike natural object recognition that it also provides no real insights. Even if object recognition research is to focus on how we use *shape* information to discriminate between objects, important sources of such information are not available if objects are depicted flat and subjects are not able to change their view by self-movement. The absence of this information, and the exact nature of the other information that is available, seems as likely to affect viewpoint constancy as the nature of the underlying object representations.

REFERENCES

Bar, M. (2001). Viewpoint dependency in visual object recognition does not necessarily imply viewer-centered representation. *Journal of Cognitive Neuroscience*, *13*, 793–799.

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. (2000) Differing views on views: Response to Hayward and Tarr (2000). *Vision Research*, *40*, 3901–3905.

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 Bulthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1506–1514.

Bulthoff, H. H., & Edelman, S. (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences USA*, *89*, 60–64

Burghund, E. D., & Marsolek, C. J. (2000) Viewpoint-invariant and viewpoint-dependent object recognition in dissociable neural subsystems. *Psychonomic Bulletin and Review*, *7*, 480–489.

Edelman, S. (1995). Representation, similarity, and the chorus of prototypes. *Minds and Machines*, *5*, 45–68.

Edelman, S., & Bulthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, *32*, 2385–2400.

Edelman, S., & Intrator, N. (2003a). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*, *27*, 73–109.

Edelman, S., & Intrator, N. (2003b). Better limited systematicity in hand than structural descriptions in the bush: A reply to Hummel. *Cognitive Science*, *27*, 331–332.

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society London B: Biological Sciences*, *269*, 1939–1947

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2002). BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron*, *34*, 161–171.

Hayward, W. G. (2003). After the viewpoint debate: Where next in object recognition? *Trends in Cognitive Sciences*, *7*, 425–427.

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., & Tarr, M. J. (2000). Differing views on views: Comments on Biederman and Bar (1999). *Vision Research*, *40*, 3895–3899.

Hummel, J. E. (2003). ''Effective systematicity'' in, ''effective systematicity'' out: A reply to Edelman and Intrator (2003). *Cognitive Science*, *27*, 327–329.

Humphrey, G. K., & Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology*, *46*, 170–190.

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica 102*, 221–245.

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.

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

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

Simons, D. J., Wang, R. X. F., & Roddenberry, D. (1998) Object recognition is mediated by extraretinal information. *Perception and Psychophysics*, *64*, 521–530.

Stankiewicz, B. J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 913–932.

Stankiewicz, B. J. (2003). Just another view. *Trends in Cognitive Sciences*, *7*, 526.

Tarr, M. J., & Bulthoff, H. H. (1995). Is human object recognition better described by geon-structural descriptions or by multiple-views? *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1494–1505.

Tarr, M. J., & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, *67*, 1–20.

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences*, *7*, 23–30.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology*, *21*, 233–282.

Vanrie, J., Beatse, E., Wagemans, J., Sunaert, S., & Van Hecke, P. (2002). Mental rotation versus invariant features in object perception from different viewpoints: An fMRI study. *Neuropsychologia*, *40*, 917–930.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, *32*, 193–254.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition*, *67*, 21–44.

Williams, P., & Tarr, M. J. (No date). *RSVP: Experimental control software for MacOs* [Online]. Available: http://www.cog.brown.edu/~tarr/RSVP/

Manuscript received 23 February 2003
Manuscript accepted 27 July 2004