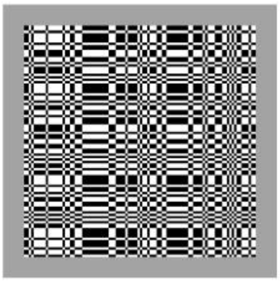
VISION RESEARCH

An International Journal for Functional Aspects of Vision



Biochemistry & Cell Biology • Molecular Biology & Genetics
Anatomy, Physiology, Pathology & Pharmacology • Optics, Accommodation & Refractive Error
Circuitry & Pathways • Psychophysics • Perception • Attention & Cognition
Computational Vision • Eye Movements & Visuomotor Control

ISSN 0042-6989 | Volume 47 | Number 16 | July 2007

Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 2164–2169
Vision Research
www.elsevier.com/locate/visres

Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces

Darren Burke *, Jessica Taubert, Talia Higman

Centre for the Integrative Study of Animal Behaviour, Macquarie University, Sydney, NSW 2109, Australia

Received 1 February 2007; received in revised form 18 April 2007

Abstract

Almost all previous studies of face recognition have found that matching the same face depicted from different viewpoints incurs both reaction time and accuracy costs. This has been interpreted as evidence that the underlying neural representations of faces are viewpoint-specific, but such a conclusion depends on the experimental data being an accurate reflection of real-world viewpoint generalisation. An equally plausible explanation for poor viewpoint generalisation in experimental situations is that important information that is normally used to generalise across views in real-world settings is not available in the experiment. Stereoscopic information about the three-dimensional structure of the face is systematically misleading in nearly all previous investigations of face recognition, since a face depicted on a computer monitor contains explicit stereoscopic information that the face is flat. The current experiment demonstrates that viewpoint costs are reduced by depicting the face with stereoscopic three-dimensionality (compared to a synoptically presented face), raising the possibility that the viewpoint costs found in face recognition experiments might be a better reflection of the information that is typically unavailable in the experimental stimuli than of the underlying neural representation of facial identity.

Keywords: Face perception; Face recognition; Viewpoint; Stereopsis

1. Introduction

In order to recognise the faces of other individuals in real world settings, we need to generalise across a range of input transformations, each of which poses serious challenges to the visual system. The fact that dealing with such transformations is challenging is evidenced by the surprisingly poor performance of subjects who are asked to match faces of unfamiliar individuals taken from slightly different camera angles in slightly different lighting conditions (Hancock, Bruce & Burton, 2000). In the general object recognition literature, the transformation that has been most studied is changes in the viewpoint from which the object is depicted, since this produces the most striking changes to the input image. Although debate continues, the vast majority of studies using novel objects have found that performance is strongly viewpoint-dependent, falling off linearly as the difference between the to-be-matched views increases (e.g., Lawson, 1999; Tarr & Pinker, 1989; reviewed by Hayward, 2003; Tarr & Cheng, 2003). This has led most researchers to conclude that the underlying neural representations by which objects are recognised are also narrowly tuned to particular previously seen views, a claim for which there is neurophysiological support from single cell recording (Logothetis & Sheinberg, 1996) and neuroimaging (Gauthier et al., 2002), studies (but see Bar, 2001; Burke, 2005; Stankiewicz, 2003, for alternative perspectives).

Although it has been of much less concern to face researchers, those studies that have examined viewpoint generalisation have also uniformly found performance costs as viewpoint differences increase (Hill, Schyns, & Akamatzu, 1997; Lee, Matsumiya, & Wilson, 2006; Newell, Chiroro, & Valentine, 1999; O'Toole et al., 1998; Troje & Bulthoff, 1996, 1998; Troje & Kersten, 1999), and they have also drawn the implication that the neural representation

* Corresponding author.
E-mail address: darren@galliform.bhs.mq.edu.au (D. Burke).

0042-6989/$ - see front matter Crown copyright © 2007 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2007.04.018

of faces must therefore be viewpoint-specific. This conclusion is supported by the existence of view-specific face-sensitive neurons in monkey inferotemporal cortex (Perrett, Hietanen, Oram, & Benson, 1991), and view-sensitive adaptation of neural activity in human lateral occipital cortex when faces are used as adapting stimuli (Grill-Spector et al., 1999), as well as evidence of viewpoint aftereffects (Fang & He, 2005), in which adaptation to a face depicted from 30 degrees to the right of directly in front, causes a front-viewed face to appear to be seen slightly from the left. Of course, the view-specific face-sensitive neurons revealed by these studies are not necessarily involved in recognising particular individuals, since viewpoint is an important piece of information (especially for socially relevant stimuli like faces) that might be encoded in its own right, and so they may not be at the heart of viewpoint-dependent performance on face recognition tasks. Somewhat more compelling, is recent evidence that face-distortion aftereffects (Webster & MacLin, 1999) are strongly tuned for viewpoint, showing poor transfer to non-adapted views (Jeffrey, Rhodes, & Busey, 2006), but again, this is only evidence of view-specific encoding of facial *identity* information if the distortion in the face (contraction or expansion of internal facial features) is registered as a change in facial identity, which it may not be.

A more fundamental issue is whether the viewpoint-dependent performance in laboratory-based recognition or matching tasks accurately reflects real-world recognition performance, since many sources of real-world information are absent in most experiments. If experimentally measured viewpoint costs (both behavioural and neurophysiological) are actually the result of using stimuli that do not contain the information that is normally used to generalise between views, then theories of face representation based on those costs may be premature. Stereopsis is an obvious source of information that is systematically misleading in all of the studies reviewed earlier showing viewpoint-dependent recognition performance. When a face is viewed binocularly on a computer monitor, there is explicit stereoscopic information provided specifying that the face is, in fact, flat. Although we are well practiced at recognising individuals despite this misleading stereoscopic information, and there is evidence that the addition of stereoscopic information does not greatly assist recognising faces depicted from a single viewpoint (Liu, Collin, & Chaudhuri, 2000; Liu & Ward, 2006; Liu, Ward, & Young, 2006), there are good reasons for believing that stereopsis might be particularly helpful for viewpoint generalisation. There is a great deal of information about the three dimensional structure of a face that is simply unavailable if it is presented from directly in front and flat, that might assist in matching the face at a novel viewpoint (the extent to which the nose protrudes, for example). There is also evidence that stereoscopic information assists viewpoint generalisation with novel objects (Bennett & Vuong, 2006; Burke, 2005), and so it would be theoretically interesting if it failed to assist viewpoint generalisation with faces.

The current experiment examined whether, despite previous evidence that three-dimensional information is unimportant in face recognition, the addition of stereoscopic information to a face might help participants to match individuals depicted from different viewpoints, a fundamental task in face recognition.

2. Methods

2.1. Subjects

Twenty students of Macquarie University and two of the authors participated in the experiment. 11 subjects (9 female) participated in the stereoscopic condition and 11 subjects (seven female) participated in the synoptic condition. Both authors served as subjects in the synoptic condition in case their familiarity with the faces improved their performance. We wished to avoid the possibility that better performance overall in the stereoscopic condition was being driven by a few good subjects.

2.2. Materials and procedure

Faces of 6 Caucasian males aged between 20 and 25 were used as test stimuli (see Fig. 1). Subjects wore a swimming cap to remove distinctive hair- and ear-based discrimination cues. Two Kodak (DX6490 4 megapixel) digital cameras were positioned 70 mm apart, at nose height, 1 meter from the subject. Each subject was photographed simultaneously by each camera from 3 different viewpoints (see Fig. 1), to create the stereopairs. The digital images were cropped, reduced to 8-bit greyscale and resized so that they were all the same height (7.8 cm). This size was chosen to preserve the disparity present in the original photographs when they were viewed through the stereoscope. Viewpoint changes meant that the images differed in width (from 4.9–6.2 cm). In the experiment, the stimuli were viewed through a Stereo Aids ScreenScope mirror stereoscope (www.stereoaids.com.au) with an effective viewing distance of 40 cm to each eye. The stereoscope provided a head- and nose-rest that minimized subjects' head movement. Experimental subjects were presented with either both the left-eye and right-eye view of the face, so that they saw all stimuli depicted with stereoscopic three dimensionality, or with the left-eye view presented to both eyes, producing a synoptic depiction of all faces (see Fig. 2).

A synoptic presentation, unlike normal binocular viewing of a face on a computer screen, contains *no* stereoscopic information, simulating a face too far away to provide disparity cues. Stimuli were presented on a Sony Trinitron CPD G520 RGB monitor, driven by a Power Macintosh G4. Each subject participated in either the Stereo or the Synoptic condition, in which all stimuli were viewed stereoscopically or synoptically. Viewing condition (Stereo or Synoptic) was manipulated between subjects because viewing the faces stereoscopically might be expected to give subjects sufficient structural information about each individual face to assist their performance in *both* conditions, had we manipulated this variable within subjects. This would potentially have reduced genuine differences between viewing conditions. Synoptic viewing was preferred to a flat depiction of the faces in order to equate the "realism" of the faces. Of course, a depiction of the faces through a stereoscope does not preserve all of the cues that are available in judging the depth in real faces (accommodation cues and the gradient of focal blur, for example – Watt, Akeley, Ernst, & Banks, 2005), and there is evidence that synoptic presentations provide better depth percepts than flat depictions (Koenderink, van Doorn, & Kappers, 1994), but we wanted to compare viewpoint generalization in a condition with full stereo cues to one with *no* stereo cues, rather than one with conflicting stereo cues. Stimulus presentation and data collection were controlled by RSVP (Williams and Tarr).

Subjects performed a successive, same–different task, in which they saw the first face for 2500 ms (always shown from front-on), which was replaced by a pattern mask (made up of randomly shuffled segments of

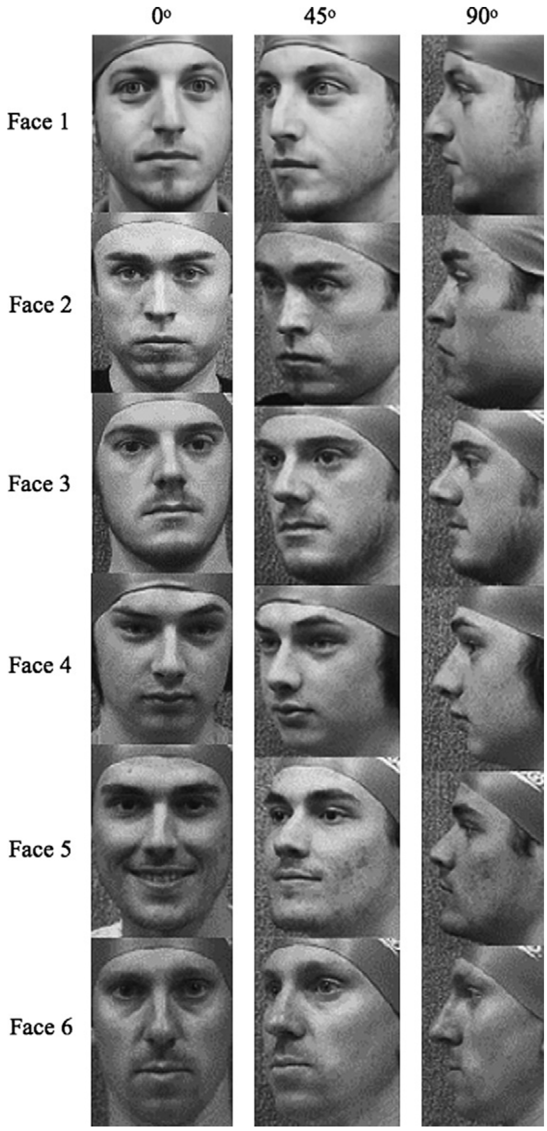


Fig. 1. The left eye views of each of the faces at each of the three viewpoints used in the experiment.

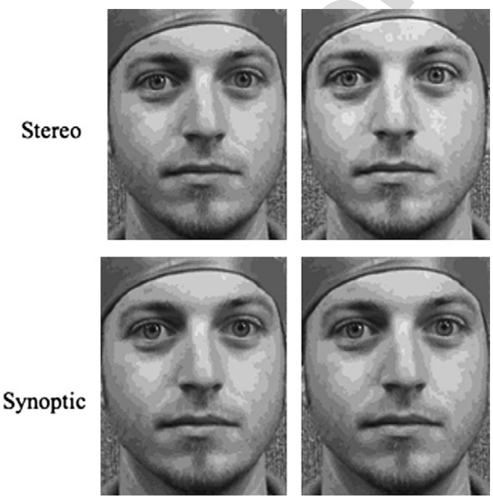


Fig. 2. Stereopairs of Face 1 at the 0° viewpoint. The stereopairs can be fused by looking through the page.

the face images) for 500 ms, and then the second face 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 face was shown from a different viewpoint or was different size, and to respond "different" only if the two presentations were of different people. In the trials of interest, the second face was of the same individual depicted at either the same viewpoint (0°), in three-quarter view (45°) or in profile (90°). In order to examine whether stereoscopic information was especially useful for generalizing across viewpoints, on half the trials the second face was also smaller than the first face (scaled to 3.9 cm tall – 25% of the size of the first face). The basic trial structure of the experiment consisted of the "same" trials in which each of the six faces was paired with itself at three viewpoints and two sizes, making 36 trials. There were three presentations of these 36 trials. There were also 108 "different" trials, which were constructed by replacing the second face in each trial with a different face, but shown from the same viewpoint and the same size as the face it replaced. The replacement procedure cycled systematically through the "different" faces. The resulting 216 trials were presented in a different random order to each subject. Subjects were presented with auditory feedback when they made an error.

3. Results

3.1. Percent correct data

Percent correct data from "same" trials were subjected to a 2 (Viewing Condition; Synoptic or Stereo) × 3 (Viewpoint; 0°, 45° or 90°) × 2 (Size; same or smaller) mixed factorial ANOVA, with Viewing Condition as a between subjects factor and Viewpoint and Size as within subjects factors. As is suggested by Fig. 3 there was a significant main effect of Viewpoint ($F_{2,40} = 56.94$, $p < 0.001$, and a significant interaction between Viewing Condition and Viewpoint ($F_{2,40} = 4.31$, $p = 0.020$), with shallower viewpoint costs in the Stereo condition. No other effects were significant (all $F < 1$).

The differences between the stereo and synoptic condition are unlikely to be due to a response bias since the mean performance difference on "different" trials perfectly matches the difference observed on same trials (see Fig. 3).

3.2. Reaction time data

Reaction time data from correct "same" trials were also subjected to a 2 (Viewing Condition; Synoptic or Stereo) × 3 (Viewpoint; 0°, 45° or 90°) × 2 (Size; same or smaller) mixed factorial ANOVA, with Viewing Condition as a between subjects factor and Viewpoint and Size as within subjects factors. In this analysis only the main effect of Viewpoint was significant ($F_{2,40} = 36.25$, $p < 0.001$), although the main effect of Size approached significance ($F_{1,20} = 3.57$, $p = 0.073$), but as is clear from Fig. 4, there is no evidence that the interaction between Viewpoint and Viewing Condition found in the Percent Correct data was due to a speed/accuracy tradeoff. No other effects were significant.

4. Discussion

The data from this experiment clearly show that the inclusion of stereoscopic information in a face helps

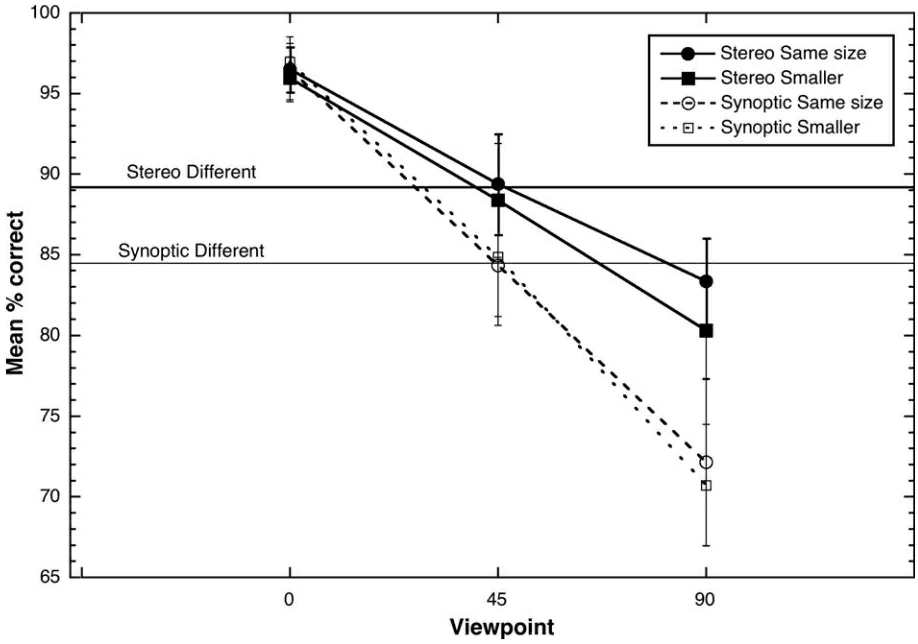


Fig. 3. Mean percent correct data as a function of viewpoint at which the second face was depicted. Viewpoint costs are shallower in the Stereo conditions than in the Synoptic conditions. Error bars represent ±1 standard error of the mean. Also plotted is the mean performance on different trials in the Stereo and Synoptic conditions.

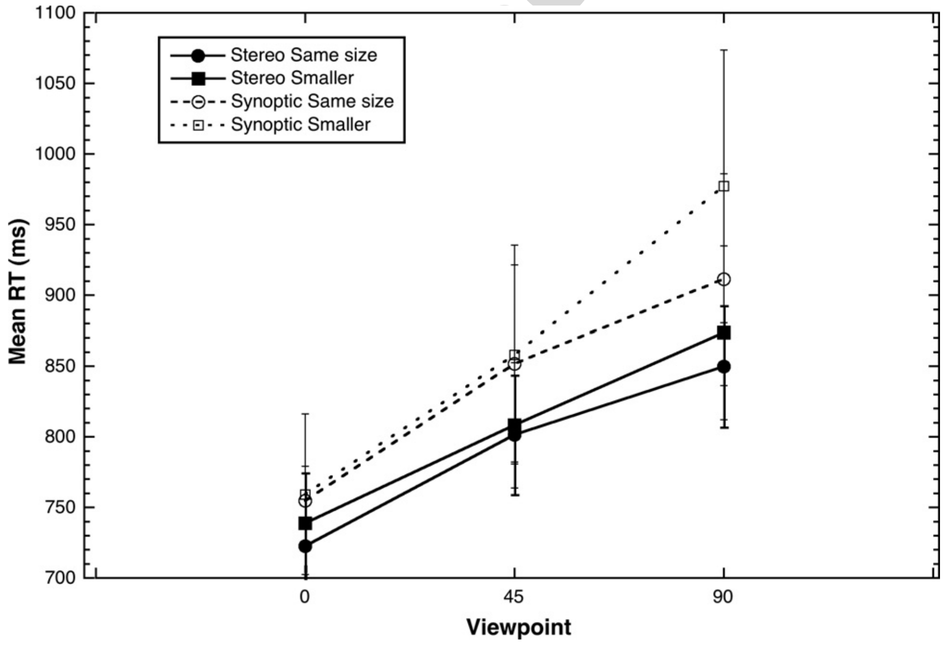


Fig. 4. Mean reaction time (RT) as a function of the viewpoint at which the second face was depicted. Error bars represent ±1 standard error of the mean.

subjects to recognise that face at new viewpoints, as reflected by the shallower viewpoint cost slopes in the percent correct data. This was true even when performance was compared to a Synoptic condition, in which the face was presented without ambiguous flat stereoscopic information. Performance in a ‘‘flat’’ condition, as is typical in face recognition experiments, might be expected to be somewhat worse than in the Synoptic condition, although this is clearly an empirical question. There was no evidence that stereoscopic information helped subjects to recognise the face at the same viewpoint (consistent with previous findings of Liu and colleagues), and no evidence that it helped to generalise across changes in size, suggesting that it is particularly helpful for viewpoint generalisation. Although possible, it is also unlikely that the additional viewpoint cues provided in the stereoscopic condition (since two slightly different views of each face were presented simultaneously in this condition) are responsible for the stereo advantage observed. First, such cues were essentially undetectable when the stereoimages were fused, and second, a similar experiment with novel objects showed no advantage for stereopairs presented side by side without a stereoscope (Burke, 2005).

As has been argued elsewhere for other kinds of objects (Burke, 2005), such data are difficult to reconcile with the idea that viewpoint-specific neural representations underlie our ability to recognise particular individuals, because although the addition of three-dimensional structural information might be expected to improve the quality of the representation at any given viewpoint, it would not obviously help to generalise *between* viewpoints if the representations were genuinely viewpoint-specific. Of course, performance in this experiment was still viewpoint dependent in all conditions (just *less* viewpoint dependent in the Stereo condition), so the data do not imply that the neural representation of facial identity is viewpoint-independent either. Rather, the data suggest that at least part of the viewpoint-dependency found in previous studies is due to using stimuli that exclude a useful source of information for generalising across viewpoints – the three-dimensional structure of the face.

If the viewpoint costs found in experiments are at least partly due to the use of impoverished stimuli rather than viewpoint-specific neural representations of facial identity, then the addition of other information about the three dimensional structure of the face that is available in everyday viewing might also be expected to reduce viewpoint cost. Consistent with this idea, viewpoint costs are also reduced by using face stimuli that are depicted in three quarter view (Hill et al., 1997), that rigidly rotate, providing motion-based structural information (Pike, Kemp, Towell N., & Phillips, 1997), or that depict individual-typical, non-rigid internal motion (Watson, Johnston, Hill, & Troje, 2005). All of these sources of information, along with stereo and a host of others, are typically available in the faces our visual system evolved to identify.

What the current results, and those like them, imply is that the visual system uses a range of different sources of information to perform complex recognition tasks. From this perspective, there is genuine theoretical progress to be made investigating what kinds of information are most useful for dealing with particular kinds of image changes – stereopsis is useful for generalising across changes in viewpoint but not across changes in size, for example. On the other hand, these data also highlight the limitations of general theories of face and object recognition formulated on the basis of the costs associated with generalising across one kind of transformation (viewpoint changes) under conditions in which much of the normally useful information is either absent or misleading.

Acknowledgments

We thank Steve Gibson (and friends) for producing the photographs used in the experiment, and two anonymous reviewers for helpful comments in the preparation of the manuscript.

References

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

Bennett, D., & Vuong, Q. C. (2006). A stereo advantage in generalising over changes in viewpoint on object recognition tasks. *Perception and Psychophysics, 68*, 1082–1093.

Burke, D. (2005). Combining disparate views of objects: Viewpoint costs are reduced by stereopsis. *Visual Cognition, 12*, 705–719.

Fang, F., & He, S. (2005). Viewer-centered object representation in the human visual system revealed by viewpoint aftereffects. *Neuron, 45*, 793–800.

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.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*, 187–203.

Hancock, P. J. B., Bruce, V., & Burton, M. A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*, 330–337.

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

Hill, H., Schyns, P. G., & Akamatzu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201–222.

Jeffrey, L., Rhodes, G., & Busey, T. (2006). View-specific coding of face shape. *Psychological Science, 17*, 501–505.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1994). On so-called paradoxical monocular stereoscopy. *Perception, 23*, 583–594.

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

Lee, Y., Matsumiya, K., & Wilson, H. R. (2006). Size-invariant, but viewpoint-dependent representation of faces. *Vision Research, 46*, 1901–1910.

Liu, C. H., & Ward, J. (2006). The use of 3D information in face recogniton. *Vision Research, 46*, 768–773.

Liu, C. H., Collin, C. A., & Chaudhuri, A. (2000). Does face recognition rely on encoding of 3-D surface? Examining the role of shape-from-shading and shape-from-stereo. *Perception, 29*, 729–743.

Liu, C. H., Ward, J., & Young, A. W. (2006). Transfer between two- and three-dimensional representations of faces. *Visual Cognition, 13*, 51–64.

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

Newell, F. N., Chiroro, P., & Valentine, T. (1999). Recognising unfamiliar faces: The effects of distinctiveness and view. *Quarterly Journal of Experimental Psychology, 52A*, 509–534.

O'Toole, A. J., Edelman, S., & Bulthoff, H. H. (1998). Stimulus specific effects in face recognition over changes in viewpoint. *Vision Research, 38*, 2351–2363.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1991). Organisation and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences, 335*, 23–31.

Pike, G., Kemp, R., Towell, N., & Phillips, K. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition, 4*, 409–437.

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

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.

Troje, N. F., & Bulthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*, 1761–1771.

Troje, N. F., & Bulthoff, H. H. (1998). How is bilateral symmetry of human faces used for recognition of novel views? *Vision Research, 38*, 78–89.

Troje, N. F., & Kersten, D. (1999). Viewpoint-dependent recognition of familiar faces. *Perception, 28*.

Watson, T. L., Johnston, A., Hill, H. C. H., & Troje, N. F. (2005). Motion as a cue for viewpoint invariance. *Visual Cogniton, 12*, 1291–1308.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision, 5*, 834–862.

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin and Review, 6*, 647–653.

Williams, P., & Tarr, M. J. RSVP: Experimental control software for MacOs [Online]. Available: http://www.cog.brown.edu/~tarr/RSVP/.