

# Learning illumination- and orientation-invariant representations of objects through temporal association

Queensland Brain Institute, University of Queensland,  
QLD, Australia, Max-Planck-Institut für biologische Kybernetik,  
Tübingen, Germany, &  
School of Human Movement Studies,  
University of Queensland, QLD, Australia

**Guy Wallis**



**Benjamin T. Backus**

SUNY State College of Optometry, Vision Sciences,  
New York, NY, USA



**Michael Langer**

Max-Planck-Institut für biologische Kybernetik, Tübingen, Germany, &  
McConnell Eng., McGill University, Montreal, Canada



**Gesche Huebner**

Justus-Liebig-Universität Gießen, Fachbereich 06,  
Psychologie und Sportwissenschaft,  
Abteilung Allgemeine Psychologie, Gießen, Germany, &  
School of Human Movement Studies,  
University of Queensland, QLD, Australia



**Heinrich Bülthoff**

Max-Planck-Institut für biologische Kybernetik,  
Tübingen, Germany



As the orientation or illumination of an object changes so does its appearance. This paper considers how observers are nonetheless able to recognize objects that have undergone such changes. In particular the paper tests the hypothesis that observers rely on temporal correlations between different object views to decide whether they are views of the same object or not. In a series of experiments subjects were shown a sequence of views representing a slowly transforming object. Testing revealed that subjects had formed object representations which were directly influenced by the temporal characteristics of the training views. In particular, introducing spurious correlations between views of different people's heads caused subjects to regard those views as being of a single person. This rapid and robust overriding of basic generalization processes supports the view that our recognition system tracks the correlated appearance of views of objects across time. Such view associations appear to allow the visual system to solve the view invariance problem without recourse to complex illumination models for extracting 3D form, or the use of the image plane transformations required to make appearance-based comparisons.

Keywords: object recognition, learning, face recognition, visual cognition

Citation: Wallis, G., Backus, B. T., Langer, M., Huebner, G., & Bülthoff, H. (2009). Learning illumination- and orientation-invariant representations of objects through temporal association. *Journal of Vision*, 9(7):6, 1–8, <http://journalofvision.org/9/7/6/>, doi:10.1167/9.7.6.

## Introduction

Classical approaches to object recognition focus on deconstructing the retinal image into cues relating to 3D shape such as depth and edge junctions (Biederman, 1987; Marr & Hildreth, 1980). An alternative model proposes that recognition is based upon image matching (Bülthoff & Edelman, 1992) and more recently, abstract feature matching (Torralba, Murphy, & Freeman, 2007; Ullman, 2006; Wallis & Bülthoff, 1999). In its simplest form, the image-based approach can be thought of as gathering

snap-shots of an object taken under varying viewing conditions. Recognition simply requires matching new images to any one of the stored images. A challenge for such a model is explaining how the system knows which views are views of a particular object. Associating object views according to the similarity of spatial characteristics can, at best, only provide limited tolerance to variations in an object's appearance. Alternatively, the visual system could associate views on the basis of their temporal proximity (Miyashita, 1993; Pitts & McCulloch, 1947; Wallis & Bülthoff, 1999). Temporal proximity is informative because successive views are likely to come from a single (possibly

transforming) object. As we turn a box in our hand, for example, it produces a stream of reproducible, temporally correlated views. Associating views in this way has the advantage that it is useful for invariance learning across all manner of naturally occurring transformations including rotation in depth, spatial shifts and in-plane rotations, size changes, illumination changes, non-rigid motion, and so on.

The literature on object recognition in humans contains evidence both for and against the importance of sequential association during learning. For objects that rotate in depth, sequential views do come to be associated with one another in a manner that aids recognition (Liu, 2007; Stone, 1998; Vuong & Tarr, 2004; Wallis & Bühlhoff, 2001). However, there have also been counter examples, suggesting that temporal association is neither necessary for view generalization (Wang, Obama, Yamashita, Sugihara, & Tanaka, 2005), nor even beneficial (Harman & Humphrey, 1999). One way of reconciling these results might be to suggest that sequential association is important for sequences in which the majority of object parts change (as is true of rotation in depth), but not in cases where views can be easily associated on the basis of shared features. Indeed it makes intuitive sense. However, we do not believe that temporal association is therefore limited to helping observers cope with rotation in depth. In this paper we report face recognition studies looking at image-plane rotation and changes of illumination, in which object (face) parts were neither gained nor lost. We conclude that the manner in which recognition is generalized across views reflects a process by which object representations are built up, in part, from associated sequential views.

## General methods

### Background

The temporal association hypothesis predicts that views of objects are associated as examples of a single object

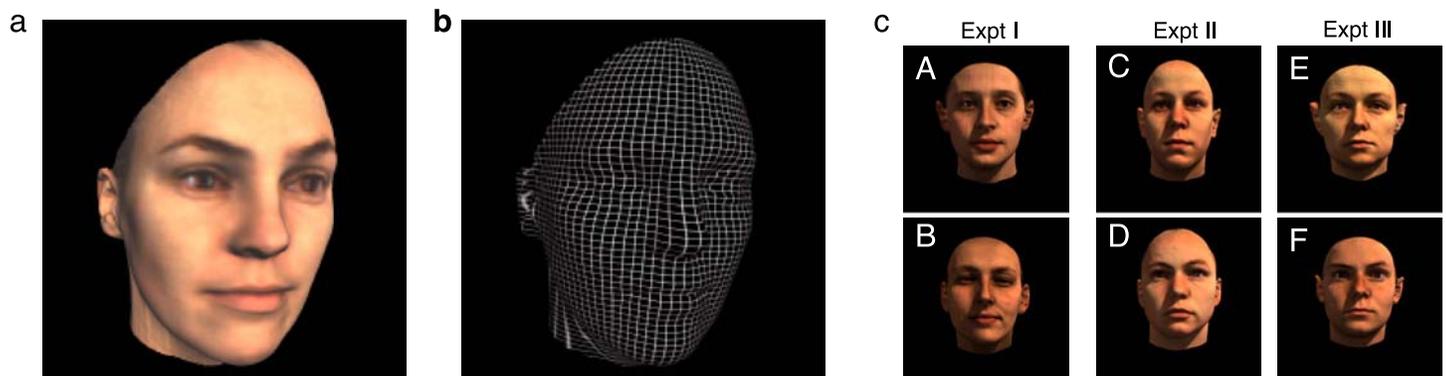


Figure 1. The faces presented during the experiment are rendered views of a three-dimensional head model. Each head consists of a) a textured surface and b) a surface mesh. c) Examples of the face pairs used in the three experiments. Each experiment used a unique set of twenty heads of this type.

simply on the basis of their being temporally proximate. In order to test this, subjects were exposed to sequences of images which altered the temporal presentation order of certain views of a person's head. The basic methodology involved displaying a head undergoing a natural change in appearance, while simultaneously undergoing a change in identity from person A to person B—see [Figures 1 and 2](#). According to the temporal association hypothesis, exposure to the consistent association of two different people across two different viewing conditions should cause the views of their heads to be regarded as belonging to the same person.

### Observers

Twenty-four participants with corrected to normal vision were tested in three separate experiments. All 24 were naive as to the purpose of the experiment and were tested in accordance with the rules and regulations of the University of Queensland's Behavioural and Social Sciences Ethical Review Committee.

### Procedure

In each experiment the participants sat 60 cm from a 24" Sony Trinitron monitor observing the projection of a 3D head model displayed centrally, and subtending an angle of approximately.

Each experiment consisted of three interleaved blocks of sequence presentation and testing. During the exposure phase, participants were presented with a total of ten heads. Each presentation consisted of the head being displayed in seven different poses at 200 ms per image—see [Figure 1](#). By presenting the images in rapid sequential order the head appeared to undergo a smooth transformation. The sequence was played back and forth for a total of 8.4 seconds. During each exposure phase, the faces were presented in pseudo-random order twice. Each subject

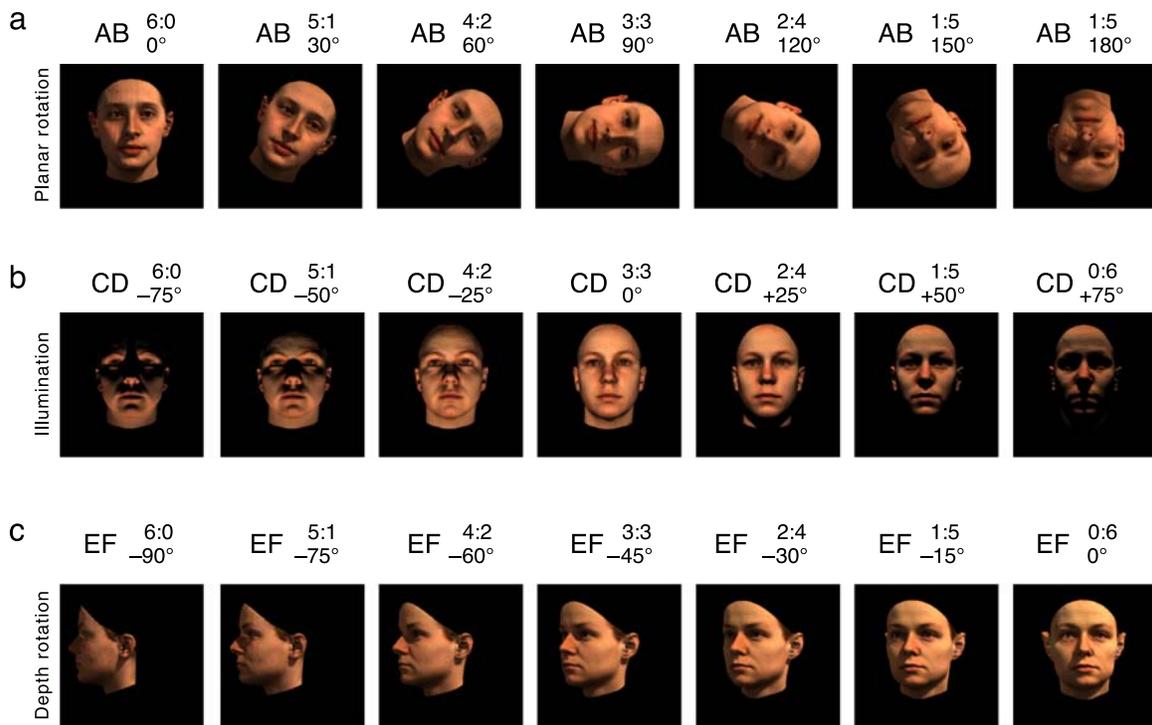


Figure 2. Exposure sequences for the three experiments. a) In [Experiment I](#), as the head is rotated in the plane, the face is also morphed from person A to person B. The proportions of the two faces are listed above the image ranging from 6:0 (indicating all head A and none of head B) in the upright view, through to all of head B and none of head A in the inverted view. b) Similarly in [Experiment II](#), illumination varies from below the mid-line to above. c) In [Experiment III](#) the head rotates in depth from profile to frontal view. In all three experiments, in the training phase subjects either saw morph sequences such as those depicted above (AB, CD and DE), or veridical transformation sequences (AA, CC and EE etc.).

saw a mixture of heads: Five involved veridical changes in appearance across the transformation and the other five involved a simultaneous morph from one person to another. The choice of which of the heads were shown in veridical form and which in morphed form was randomized for each subject.

During the testing phase of each block the same ten heads seen during the exposure phase were presented in a sequential, same/different task. The two face views displayed on a given trial corresponded to two of the extreme poses from the original animation sequences, e.g. frontal and profile view of the depth-rotated heads or upright versus inverted for the picture-plane-rotated heads. Each image appeared for 200 ms with a 500 ms inter-stimulus interval. In each test phase all distractor and match views were tested three times for a total of sixty test trials per block.

## Stimuli

The stimuli were prepared using a subset of the 3D Head Models Database from the Max Planck Institute in Tübingen. The heads were originally scanned using a

Cyberware 3D laser scanner which samples texture and shape information on a regular cylindrical grid, with a resolution of 0.8 degrees horizontally and 0.6 mm vertically. For the purposes of this experiment the heads of sixty female volunteers were chosen from amongst a set of 106. The heads were selected in pairs that ranked highly on a confusion matrix of face pairs selected from the 106 faces. This matrix was generated from an earlier discrimination task run on a separate group of 10 subjects. The purpose of this manipulation was to pair heads that were regarded as perceptually similar, and therefore easily confused. Ten pairs of heads were allocated to each of the three experiments. The head morphing technology is described elsewhere (Vetter, 1998).

## Experiment I

### Background

In this first experiment observers viewed sequences of faces undergoing rotation in the image plane, see [Figure 2a](#). This type of rotation is of particular interest because it

maintains the presence of all image features useful to recognition. Nevertheless humans are relatively poor at recognizing faces upside-down (Yin, 1969) which speaks against humans being able to perform even basic image transformations well. (The interested reader can confirm this by trying to identify faces of famous people from a newspaper which has been turned upside down).

## Method

During the initial exposure phase, subjects saw ten sequences, half of which depicted valid transformations and the other half invalid (i.e. incorporating a switch in identity). To test the effect of the exposure, participants performed a same-different discrimination task on two views that depicted one completely upright and one completely upside-down face. The temporal association hypothesis predicts that views seen in close temporal proximity will be regarded as views of the same face. Hence learning should worsen performance specifically for trained non-veridical view pairings. Unlike some previous studies (Wallis, 2002; Wallis & Bühlhoff, 2001), this study uses a balanced design in which half of the training pairs are actually veridical. This is helpful in establishing baseline performance for novel and trained veridical faces. Our measure is the differential effect of learning on veridical versus non-veridical sequences and

therefore whether there is a reliable interaction term in a 2-way ANOVA.

## Results

The temporal association hypothesis does not make specific predictions about main effects, just whether the interaction term is significant. To establish whether such an interaction occurred in this experiment a repeated-measures ANOVA was conducted with same/different and seen/unseen as independent variables, and percent correct performance as dependent variable. No significant main effects emerged, but there was a significant interaction  $F(1,23) = 4.8$ ,  $MSe = 146.7$ ,  $p = 0.039$ , Cohen's  $d = 0.54$ . The interaction revealed that exposure had produced a relative increase in performance for SAME trials, i.e. for those stimuli which had been seen in unmorphed, valid transform sequences. At the same time there was a relative decrease for DIFF trials, i.e. for those stimuli which had been seen in morphing, invalid sequences. This difference is depicted graphically in Figure 3. To investigate this effect further a second ANOVA was carried out focused on DIFF trial performance with and without training. The analysis revealed a significant effect of training  $F(1,23) = 7.49$ ,  $MSe = 142.8$ ,  $p = 0.012$ , Cohen's  $d = 0.44$ , indicating that subjects were more likely to regard two different faces as being the same if they had appeared in

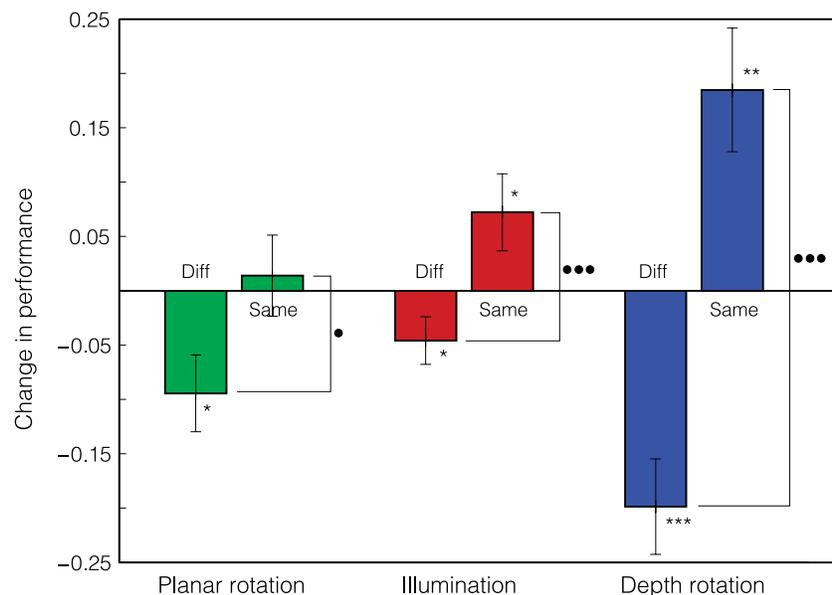


Figure 3. Results from the three experiments. In each case recognition performance is plotted for same (Same) and different (Diff) trials. Bars represent the difference in performance produced by training. A negative value indicates worse performance with prior exposure and a positive value better performance. Clearly, exposure to faces that differed (i.e. which morphed), reduced recognition performance. The decrement in performance was significant in each case. Exposure also led to generally improved performance on same trials, although the effect of exposure was only significant for depth rotation. (\* Effect of training on different and same trials, • Overall effect of training, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ). Error bars represent standard error of the mean.

one of the invalid association sequences. An analysis of SAME trials revealed an increase in performance as a result of training, but it was not statistically significant  $F(1,23) = 0.144$ ,  $MSe = 160.3$ ,  $p = 0.707$ , n.s.

## Experiment II

### Background

The second experiment followed the same logic as the first but now focused on the task of recognition across illumination changes. This too can be a difficult task and artificial vision systems often attempt to estimate a direction of illumination as part of shape constancy computations for object recognition. In reduced cue environments humans will likewise assume a direction of illumination, often that light is coming from above (Kleffner & Ramachandran, 1992)—a hard-wired assumption shared with other animals (Hershberger, 1970). Although humans are undoubtedly able to perform recognition across mild illumination changes and for moderate illumination angles (Moses, Ullman, & Edelman, 1996), generalizing facial form under highly directional illumination is not easy (Tarr, Georgiades, & Jackson, 2008).

### Method

In this experiment faces underwent a change in illumination corresponding to rotating the light source about the vertical mid-line of the face, see Figure 2b. Subjects were then tested on single frames in which two faces appeared, one illuminated from above and the other from below, both in unmorphed form.

### Results

A repeated-measures ANOVA was conducted on their discrimination performance, and again, revealed no significant main effects. There was, however, a significant interaction  $F(1,23) = 15.955$ ,  $MSe = 52.4$ ,  $p < 0.001$ , Cohen's  $d = 0.67$ . As in Experiment I the interaction was indicative of a relative increase in performance on SAME trials and a relative decrease on DIFF trials—see Figure 3. A second ANOVA was carried out focusing on DIFF trial performance as a function of training. The analysis once again revealed a significant effect  $F(1,23) = 4.54$ ,  $MSe = 55.5$ ,  $p = 0.044$ , Cohen's  $d = 0.29$ , indicating that subjects were more likely to regard views of two different people as being of the same person if those views had appeared together in one of the training sequences. An

analysis of SAME trials revealed a marginally significant increase in performance as a result of training  $F(1,23) = 4.325$ ,  $MSe = 144.8$ ,  $p = 0.049$ , Cohen's  $d = 0.38$ .

## Experiment III

### Background

In the final experiment the aim was to draw a link between this study and earlier work which has looked at rotation in depth, but using the revised training protocols (Wallis & Bülthoff, 2001).

### Method

During the exposure phase, subjects observed heads rotating in depth from left to right profile. Half of the sequences depicted veridical images associated with changes in view of the individual; in the other half the identity of the person changed smoothly from profile to frontal view. Recognition across viewpoints is not a trivial task for humans but they have been shown to perform at well above chance on faces in the database used in these experiments (Troje & Bülthoff, 1996).

### Results

In this case ANOVA did uncover a significant main effect  $F(1,23) = 5.262$ ,  $MSe = 508.1$ ,  $p = 0.031$ , Cohen's  $d = 0.50$  due to overall performance on DIFF trials falling significantly below that of SAME trials. The effect was largely driven by a very sharp fall in DIFF trial performance for trained stimuli which fell to just 45% correct. There was also a significant interaction  $F(1,23) = 22.34$ ,  $MSe = 394.7$ ,  $p < 0.001$ , Cohen's  $d = 1.11$ , due to the fact that training produced a relative increase in performance for valid view pairings, as well as the large reduction for invalid pairings—see Figure 3. An analysis of DIFF trial performance revealed a significant effect of training  $F(1,23) = 21.276$ ,  $MSe = 222.5$ ,  $p < 0.001$ , Cohen's  $d = 0.42$ , indicating that subjects were less able to discriminate faces paired together during training. Analysis of SAME trials revealed an overall increase in performance which was also statistically significant  $F(1,23) = 11.01$ ,  $MSe = 372.0$ ,  $p = 0.003$ .

### Summary of results

Across all three experiments exposure to both valid and invalid training sequences reliably increased the likelihood of subjects treating the images contained within

them as belonging to a single person. In valid sequence presentations this raised performance slightly. This increase was reliable across all three experiments but most pronounced in the illumination and depth rotation experiments. The “valid” increase, although always present, was small. This may well be due to the fact that normal performance in the three tasks is close to as good as it can be, and further training can only produce relatively small increases in performance. In contrast, the ‘unlearning’ caused by exposure to invalid transform sequences had a dramatic and statistically reliable effect in all three experiments. It is worth noting that exposure to a sequence did not simply serve to confuse subjects, as their performance on trained versus untrained objects never differed—neither systematically nor statistically reliably—in any of the three experiments.

At the conclusion of the experiments the subjects were debriefed to ascertain whether they were aware that the image sequences contained morphed images as well as veridical ones. The majority, 20 of the 24, said they were not. The data from the four subjects that did report seeing morphing taking place were later reanalyzed to see if this insight affected their results. In practice it did lead to smaller effects in two subjects, but data from the other two were indistinguishable from those that reported no awareness of morphing. When shown all 20 training sequences (10 veridical and 10 morphs), none of the four subjects were able to detect all of the morph sequences correctly. All four subjects actually reported perceiving morphs in several of the veridical sequences. This is consistent with the final experiment of one of our earlier studies which tested sequences rather than single images (Wallis & Bühlhoff, 2001).

## Discussion

In a series of three experiments new evidence has been presented extending the role of temporal association-based view-invariance learning. In a relatively short period subjects were seen to form representations of the faces shown which were directly influenced by exposure to both real and imaginary object transform sequences.

In our study, faces of close physical resemblance were deliberately paired and transformations were chosen for which recognition is notoriously difficult, namely changes in illumination, planar orientation, and rotation in depth. It was necessary to choose a task of sufficient difficulty to garner learning effects over the very short exposure periods used in these studies. After all, these studies aim to have subjects unlearn in a matter of tens of minutes, generalization abilities acquired over several decades. That is not to say that effects cannot be obtained with faces which are initially more easily distinguished, simply that it will take longer. In a lengthier study looking at categorization Preminger, Sagi, and Tsodyks (2007) were

able to redraw category boundaries by presenting cross category morphs over a period of several days.

For the purposes of these experiments the focus has been on face recognition, but we would argue that the effects are general and apply equally to other object categories. One reason for thinking this is that although the original description of temporal association-based learning concerned faces (Wallis & Bühlhoff, 2001), similar effects have since been reported for a range of novel object classes both in humans (Liu, 2007; Vuong & Tarr, 2004), and monkeys (Cox, Meier, Oertelt, & DiCarlo, 2005)—where it applied to generalizing across image displacements. The reason for choosing to study faces here is that although most people are experts in the task of face recognition, the task itself is not easy. Faces change their shape as they rotate (profile, frontal view), they self-occlude (nose, chin), they are non-rigid (smile, frown), and very similar distractors exist (other faces). It is also relatively easy to transform smoothly from one individual to another since the same major physical characteristics are present in all faces.

Head morphing was used in these experiments to enhance the spatio-temporal smoothness of the transform sequences. One concern that arises is whether the effects reported here were influenced by the morphing process. In practice, earlier studies of depth rotation (using the same face database) revealed that morphed images are neither necessary (Wallis, 2002), nor sufficient (Wallis & Bühlhoff, 2001) for image association to occur. So why use them at all? The most extreme interpretation of the temporal association hypothesis would argue that they are not necessary and that arbitrary images can be regarded as views of a single object simply by being presented in regular, temporal proximity. Indeed, evidence from prolonged studies in monkeys suggests that arbitrary image associations can be formed by single neurons trained in this way (Miyashita, 1993). Nonetheless, neural models of temporal associative learning predict that totally arbitrary associations will take a very long time to achieve, and that incremental changes in appearance across time are more likely to achieve measurable changes in neural selectivity and hence discrimination performance (Wallis, 1998). Certainly, most studies of exposure to smooth versus random sequences suggest that smooth sequences facilitate learning (Lawson, Humphreys, & Watson, 1994; Wallis & Bühlhoff, 2001). However, while this may seem intuitive, it should be noted that at least one study has found the opposite to be true (Harman & Humphrey, 1999).

An interesting question for future research is whether an awareness of the morphing—particularly if it is visually obvious—affects the learning. One might speculate that if an object is visibly transforming from one object to another, the visual system would choose not to regard the endpoints as views of the same object. Alternatively, the system may not be so sophisticated. As reported in the first experiment, at least four of our subjects were ‘suspicious’

that morphing was taking place, but were poor at distinguishing sequences containing morphs from veridical ones. As very few real objects change their identity over time, a learning module that associates sequentially presented views may simply make blind associations in the expectation that they must be valid. Certainly the early models of temporal association learning envisaged a ‘mindless’ process at work (Földiák, 1991; Wallis, 1998; Wallis & Rolls, 1997). One consequence of such a model is that it might explain the need for separate ‘where’ and ‘what’ streams. Whereas temporal association is extremely beneficial for a ‘what’ stream, it would be a great hindrance to a ‘where’ stream, which would be best served by focussing on purely spatial association (Wallis & Rolls, 1997).

One common criticism of the temporal association learning hypothesis is that recognition of novel objects is possible without prior exposure to smooth transforms of that object. In one study on monkeys, Wang et al. (2005) demonstrated good view generalization to a novel stimulus set without prior exposure to smooth transformations. That said, just as in earlier studies of view generalization in humans (Bülthoff & Edelman, 1992; Tarr & Bülthoff, 1995), performance generally dropped as a function of discrepancy of the two test views, and the monkeys failed to generalize across rotation angles beyond 60. It is precisely here that temporal association would come in to play, allowing recognition across large changes in orientation. What is also important to remember is that image-based recognition is almost certainly done not at the level of images, but rather small, abstract features. A neuron learning the invariant properties of a small constituent feature not only learns transforms relevant to that object, but also for all objects which contain that feature. This can explain why humans are often able to recognize a novel object from quite disparate views. The only requirement is that the novel object contains elements which have been observed undergoing transformations within other, familiar objects. Exposure to the novel object itself is unnecessary (Wallis & Bülthoff, 1999).

The results reported here are aligned with the view that temporal association is used to associate arbitrary views into a view set or ‘aspect graph’ (Koenderink & van Doorn, 1979) which constitutes a sufficient set of views to recognize an object from a wide range of angles, distances, retinal locations and illuminations. Other researchers have taken sensitivity to temporal characteristics of learning to mean that sequences themselves form part of an object’s description (Hill & Johnston, 2001; Stone, 1998; Vuong & Tarr, 2004). This is different from the view association model described here. In this experiment, single test images were used rather than the sequences of images seen during the exposure phase. In other words, the results here suggest that sequences are not necessary to unlock the representations, even though it was sequences of images that served to build those representations in the first place. More work will be

required to determine if and how the power of time-based view association complements or interacts with a motion/sequence-based representation of objects.

Taken as a whole the results establish the temporal association mechanism as a general purpose heuristic for learning to generalize object appearance across any changes in appearance, most especially smooth transformations, the like of which we encounter every day.

## Acknowledgments

We are grateful to Thomas Vetter and Volker Blanz who pioneered much of the head morphing technology used in this paper, and to Niko Troje for scanning and preparing the head models. We are also grateful to Stefanie Ringlehan, Kellie Swann and David Butler for help gathering data reported in this paper and for earlier pilot experiments.

This research was supported by the Australian Research Council Grant DP0343522, Human Frontiers Program Grant RGP 03/2006 and by the Max Planck Society.

Commercial relationships: none.

Corresponding author: A/Prof. Guy Wallis.

Email: gwallis@hms.uq.edu.au.

Address: Perception and Motor Systems Laboratory, School of Human Movement Studies, University of Queensland, QLD 4072, Australia.

## References

- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147. [[PubMed](#)]
- Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 60–64. [[PubMed](#)] [[Article](#)]
- Cox, D., Meier, P., Oertelt, N., & DiCarlo, J. (2005). Breaking position invariant object recognition. *Nature Neuroscience*, 8, 1145–1147. [[PubMed](#)]
- Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3, 194–200.
- Harman, K., & Humphrey, G. (1999). Encoding ‘regular’ and ‘random’ sequences of views of novel three-dimensional objects. *Perception*, 28, 601–615. [[PubMed](#)]
- Hershberger, W. (1970). Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below. *Journal of Comparative and Physiological Psychology*, 73, 407–411. [[PubMed](#)]

- Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*, 880–885. [PubMed]
- Kleffner, D., & Ramachandran, V. (1992). On the perception of shape from shading. *Perception and Psychophysics*, *52*, 18–36. [PubMed]
- Koenderink, J. J., & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics*, *32*, 211–216. [PubMed]
- Lawson, R., Humphreys, G., & Watson, D. (1994). Object recognition under sequential viewing conditions: Evidence for view-point specific recognition procedures. *Perception*, *23*, 595–614. [PubMed]
- Liu, T. (2007). Learning sequence of views of three-dimensional objects: The effect of temporal coherence on object memory. *Perception*, *36*, 1320–1333.
- Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Science*, *207*, 187–217. [PubMed]
- Miyashita, Y. (1993). Inferior temporal cortex: Where visual perception meets memory. *Annual Review of Neuroscience*, *16*, 245–263. [PubMed]
- Moses, Y., Ullman, S., & Edelman, S. (1996). Generalization to novel images in upright and inverted faces. *Perception*, *25*, 443–462. [PubMed]
- Pitts, W., & McCulloch, W. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, *9*, 127–147.
- Preminger, S., Sagi, D., & Tsodyks, M. (2007). The effects of perceptual history on memory of visual objects. *Vision Research*, *36*, 965–973. [PubMed]
- Stone, J. (1998). Object recognition using spatio-temporal signatures. *Vision Research*, *38*, 947–951.
- Tarr, M., & Bülthoff, 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., Georghiades, A., & Jackson, C. (2008). Identifying faces across variations in lighting: Psychophysics and computation. *ACM Transactions in Applied Perception*.
- Torralba, A., Murphy, K. P., & Freeman, W. T. (2007). Sharing visual features for multiclass and multiview object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*, 854–869. [PubMed]
- Troje, N., & Bülthoff, H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, *36*, 1761–1771. [PubMed]
- Ullman, S. (2006). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*, 58–64. [PubMed]
- Vetter, T. (1998). Synthesis of novel views from a single face image. *International Journal of Computer Vision*, *28*, 103–116.
- Vuong, Q., & Tarr, M. (2004). Rotation direction affects object recognition. *Vision Research*, *44*, 1717–1730. [PubMed]
- Wallis, G. (1998). Spatio-temporal influences at the neural level of object recognition. *Network: Computation in Neural Systems*, *9*, 265–278. [PubMed]
- Wallis, G. (2002). The role of object motion in forging long-term representations of objects. *Visual Cognition*, *9*, 233–247.
- Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences*, *3*, 22–31. [PubMed]
- Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 4800–4804. [PubMed] [Article]
- Wallis, G., & Rolls, E. (1997). A model of invariant object recognition in the visual system. *Progress in Neurobiology*, *51*, 167–194.
- Wang, G., Obama, S., Yamashita, W., Sugihara, T., & Tanaka, K. (2005). Prior experience of rotation is recognizing objects seen from different angles. *Nature Neuroscience*, *8*, 1768–1775. [PubMed]
- Yin, R. (1969). Looking at upside down faces. *Journal of Experimental Psychology*, *81*, 141–145.