nature neuroscience

'Breaking' position-invariant object recognition

David D Cox, Philip Meier, Nadja Oertelt & James J DiCarlo

While it is often assumed that objects can be recognized irrespective of where they fall on the retina, little is known about the mechanisms underlying this ability. By exposing human subjects to an altered world where some objects systematically changed identity during the transient blindness that accompanies eye movements, we induced predictable object confusions across retinal positions, effectively 'breaking' position invariance. Thus, position invariance is not a rigid property of vision but is constantly adapting to the statistics of the environment.

Any given object can cast an essentially infinite number of different images on the retina, owing to variations in position, scale, view, lighting and a host of other factors. Nonetheless, humans effortlessly recognize familiar objects in a manner that is largely invariant to these transformations. The ability to identify objects in spite of these transforms is central to human visual object recognition, yet the neural mechanisms that achieve this feat are poorly understood, and transform-tolerant recognition remains a major stumbling block in the development of artificial vision systems. Even for variations in the position of an image on the retina, arguably the simplest transform that the visual system must discount, little is known about how invariance is achieved.

Several authors have proposed that one solution to the invariance problem is to learn representations through experience with the spatiotemporal statistics of the natural visual world[1–4]. Visual features that covary across short time intervals are, on average, more likely to

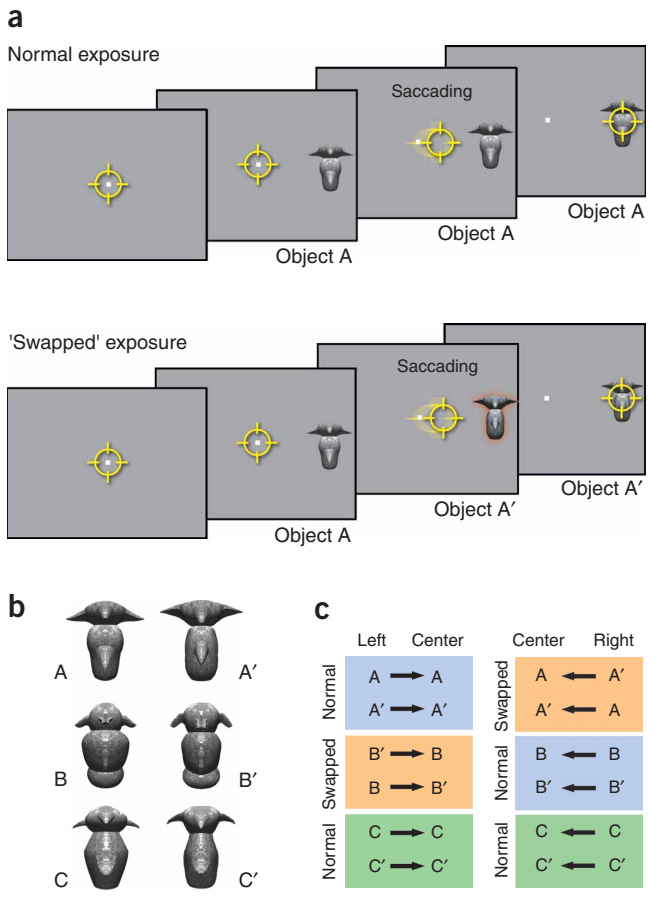

Figure 1 Experiment 1 and 2 design. Twelve naive subjects participated in each experiment and provided informed consent in accordance with the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects. (**a**) During the 'exposure' phase of each experiment, subjects received two different types of exposure trials randomly interleaved. In all trials, subjects started a trial by fixating on a point, and then an object appeared in the periphery (6° to the left or right, randomly). Subjects spontaneously saccaded to the object and were required to decide if this object was the same object as in the preceding trial. In 'normal exposure' trials, the object identity did not change, so the same object was presented to both the peripheral retina (pre-saccade) and the central retina (post-saccade). In 'swapped' exposure trials, unknown to subjects, one object was swapped for a different object in mid-saccade, such that one object was presented to the peripheral retina pre-saccade, and a different object was presented to the central retina post-saccade. (**b**) The objects used in this experiment were modified versions of the publicly available 'greeble' stimuli (**Supplementary Methods**) and were arranged in three pairs, with the differences within pair (for example, A and A′) being qualitatively smaller than the differences between pairs (for example, A and B). Objects were chosen to be relatively natural but unfamiliar to the subject. (**c**) A schematic representation of the twelve exposure trial types for one subject. All such exposure trials occurred equally often (pseudorandomly selected). Thus, each subject received an equal number of presentations of all objects in each retinal location. The letter on one side of the arrow indicates the peripherally presented object (either on the right or left), with the arrow indicating the object identity before (arrow tail) and after (arrowhead) the saccade. For all subjects, one object pair was swapped on the right but was normal on the left (top row), one pair was normal on the right but swapped on the left (middle row) and one pair was not swapped on either side (bottom row). Subjects were tested in two sets of six, with each set of six counterbalancing across all possible assignments of the three object pairs to each of these three roles. Blue panels indicate trials in which objects were not swapped; orange indicates 'swapped' trials and green indicates trials with object pairs that were not swapped on either side.

McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. Correspondence should be addressed to J.D. (dicarlo@mit.edu).

Published online 7 August 2005; doi:10.1038/nn1519

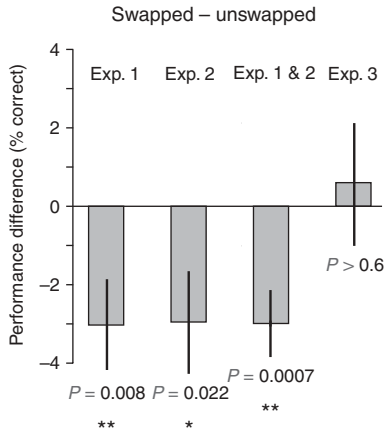

Figure 2 Results. In testing after exposure, subjects in experiment 1 (2 d of exposure; $n = 12$) and experiment 2 (1 d of exposure; $n = 12$) confused objects significantly more often across retinal positions where they had been swapped during the exposure phase (orange panels in **Fig. 1c**) than across positions where the same objects behaved normally during exposure ('unswapped'; blue panels in **Fig. 1c**). These effects were not significantly different for trials where the correct answer was the 'same' and trials where the correct answer was 'different' in either experiment 1 or 2 ($P > 0.4$ two-tailed paired *t*-tests). Subjects in experiment 3 (replay experiment; $n = 12$) who received retinal exposure matched to subjects in experiment 1 did not show a significant effect. Bars show effect magnitudes and standard errors for experiment 1 (2 d of exposure), experiment 2 (1 d of exposure), data from experiments 1 and 2 pooled together, and experiment 3 (replay). Mean performance with the control objects (green panels in **Fig. 1c**) was 74%, 72% and 78% in experiments 1, 2 and 3, respectively, and was not significantly different across the three experiments ($P > 0.1$, one-way ANOVA).

correspond to different images of the same object than to different objects, and thus one might gradually build up invariant representations by associating patterns of neural activity produced by successive retinal images of an object. While some transformations of an object's retinal image are played out smoothly across time (for example, scale and pose), changes of an object's retinal position often occur discontinuously as a result of rapid eye movements that sample the visual scene (saccades). A possible strategy, then, for building position-invariant object representations is to associate neural activity patterns across saccades, preferably taking into account the direction and magnitude of the saccade.

If correct position invariance is created through experience with the statistical properties of the visual world, it might be possible to create unnatural or 'incorrect' invariances by manipulating those statistics. In particular, if objects consistently changed their identity as a function of retinal position, then the visual system might incorrectly associate the neural representations of different objects at different positions into a single object representation. The resulting representation would be activated by one object at one retinal position and another object at another position, and thus the two objects would be perceived as being the same object at different positions.

In the present study, we engineered such a situation, taking advantage of the fact that humans are effectively blind during the short time it takes to complete a saccade[5,6]. By monitoring eye position in real time, we were able to present one object to a subject's peripheral retina that was replaced by a particular different object in mid-saccade when the subject attempted to foveate it. None of the subjects reported being aware that objects were being swapped, despite being asked in a post-session debriefing whether they had seen objects change or appear otherwise unusual. After a brief period of exposure to these altered spatiotemporal statistics (240–400 altered exposures in Experiment 1, and 120–180 altered exposures in Experiment 2), we used a same-different task to probe the subject's representations of these objects across changes in position. The layout of Experiments 1 and 2 is described in **Figure 1** and in **Supplementary Methods**.

In both experiments, subjects significantly more often confused object pairs when they were tested across the retinal positions where those particular objects had been swapped during the exposure phase than in tests across positions where the same objects had not been swapped ($P = 0.0082$ in experiment 1, $P = 0.022$ in experiment 2; $P = 0.0007$, both experiments pooled; one-tailed paired *t*-test; **Fig. 2**). That is, for previously swapped objects, subjects were more likely to perceive different objects at two retinal positions as the same object and perceive the same object at two positions as different objects.

These results show that confusions in invariant visual object processing occur after relatively brief exposure (<1 h, total) to altered spatiotemporal statistics across saccades, even though subjects were unaware of this change. Moreover, the confusions are predictable in that they are what is expected if the visual system assumes that object identity is stable across the short time interval of a saccade. Although the magnitude of the observed effect is not large, and we have shown it only for relatively similar objects, it should be borne in mind that the anomalous exposure provided represents a tiny fraction of each subject's lifetime experience with an unaltered, real-world visual environment. The ability to significantly shift object representations at all suggests that position-invariant visual object recognition is modifiable in adults, and it points to possible mechanisms by which sets of invariant features might be acquired, especially during early visual learning.

To test whether the observed effect depends critically on the execution of active eye movements, as opposed to spatiotemporal experience alone, we ran a third set of experiments (experiment 3) with twelve subjects with retinal experience matched to the subjects in experiment 1, but without saccades. These subjects maintained fixation throughout each trial during the exposure phase, and the retinal positions and timing of object exposure was 'replayed', trial-by-trial, from the spatiotemporal retinal experience generated by their counterpart subject in experiment 1. The testing phase was identical to experiments 1 and 2. Subjects in experiment 3 showed no effect of anomalous spatiotemporal experience ($P > 0.6$; one-tailed paired *t*-test, **Fig. 2**), suggesting that anomalous experience across saccades may be necessary to produce later confusions in invariant object processing.

Although these results show that specific alterations in object spatiotemporal experience can alter position-invariant recognition with test objects in the direction predicted by theory, we wondered if such anomalous experience might also produce more general deficits in recognition performance with those test objects. To examine this, we compared recognition performance of test objects across positions where those objects had behaved normally ('unswapped' conditions) with recognition of control objects (which were never swapped in either position). Although both experiments showed a trend toward reduced performance with objects whose spatiotemporal statistics had been altered (**Supplementary Fig. 1**), no significant difference was found in either experiment (experiment 1: $P = 0.48$; experiment 2: $P = 0.094$, two-tailed paired *t*-tests).

Like some recent perceptual learning studies, this study shows that visual processing can be altered by visual statistics that do not reach awareness[7]. However, in contrast to standard perceptual learning procedures in which subjects improve on some sensory task over the course of many training sessions[8], here, performance is impaired in a predictable way by brief exposure that runs counter to the subject's past visual experience. This resembles other long-term perceptual adapta-

tion effects, such as the McCollough effect and prism adaptation and, like these effects, might represent an ongoing process to adapt to the environment and keep perception veridical[9].

While adult transform-invariant object recognition is, for the most part, automatic and robust[10], this finding adds to a growing body of research suggesting that such invariance may ultimately depend upon experience[11–14]. More broadly, this finding supports the developing belief that visual representations in the brain are plastic and largely a product of the visual environment[15]. Within this context, invariant object representations are not rigid and finalized, but are continually evolving entities, ready to adapt to changes in the environment.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We would like to thank B. Balas, N. Kanwisher and P. Sinha for their helpful comments on earlier versions of this work and J. Deutsch for technical support. This work was supported by the US National Eye Institute (NIH-R01-EY014970) and the Pew Charitable Trusts (PEW UCSF 2893sc). D.D.C. is supported by a National Defense Science and Engineering Graduate Fellowship. N.O. was supported by the Paul E. Gray Memorial Undergraduate fund.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 17 May; accepted 15 July 2005
Published online at http://www.nature.com/natureneuroscience/

1. Foldiak, P. *Neural Comput.* **3**, 194–200 (1991).
2. Wallis, G. & Rolls, E.T. *Prog. Neurobiol.* **51**, 167–194 (1997).
3. Wiskott, L. & Sejnowski, T.J. *Neural Comput.* **14**, 715–770 (2002).
4. Edelman, S. & Intrator, N. *Cogn. Sci.* **27**, 73–109 (2003).
5. Ross, J., Morrone, M.C., Goldberg, M.E. & Burr, D.C. *Trends Neurosci.* **24**, 113–121 (2001).
6. McConkie, G.W. & Currie, C.B. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 563–581 (1996).
7. Watanabe, T., Nanez, J.E. & Sasaki, Y. *Nature* **413**, 844–848 (2001).
8. Karni, A. & Sagi, D. *Nature* **365**, 250–252 (1993).
9. Bedford, F. *Trends Cogn. Sci.* **3**, 4–12 (1999).
10. Biederman, I. & Bar, M. *Vision Res.* **39**, 2885–2899 (1999).
11. Dill, M. & Fahle, M. *Percept. Psychophys.* **60**, 65–81 (1998).
12. Nazir, T.A. & O'Regan, J.K. *Spat. Vis.* **5**, 81–100 (1990).
13. Dill, M. & Edelman, S. *Perception* **30**, 707–724 (2001).
14. Wallis, G. & Bülthoff, H.H. *Proc. Natl. Acad. Sci. USA* **98**, 4800–4804 (2001).
15. Simoncelli, E.P. & Olshausen, B.A. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).