dysfunction that characterizes late-onset NIDDM[27], especially if it is central in regulating gene expression in the pancreatic β-cell, as suggested by the association of mutations in the HNF-4α gene with MODY. Furthermore, the similarity between HNF-4α and ligand-dependent transcription factors raises the possibility that HNF-4α and the genes it regulates respond to an unidentified ligand. The identification of this ligand could lead to new treatments for diabetes. □

Methods

Isolation and partial sequence of the human HNF-4α gene. Three PAC clones (114E13, 130B8 and 207N8) containing the human HNF-4α gene were isolated from a library (Genome Systems) by screening DNA pools by PCR with the primers HNF4-1 (5′-CACCTGGTGATCACGTGGTC-3′) and HNF4-2 (5′-GTAAGGCTCAAGTCATCTCC-3′). The partial sequence of the gene was determined after amplifying PAC 114E13 and genomic DNA with specific primers whose sequences were based on the human cDNA[15,28] and selected using the exon–intron organization of the mouse HNF-4α gene[29] as a guide. PCR products were sequenced using an AmpliTaq FS Dye Terminator Cycle Sequencing Kit (Applied Biosystems) and an ABI Prism 377 DNA Sequencer. This gene consists of 11 exons, with the introns being located in the same positions as in the mouse gene[29]. Alternative splicing generates a family of mRNAs, HNF-4α1, 2 and 4, the latter two of which contain inserts of 30 and 90 nucleotides, respectively[15,28,29]. The sequence of exon 1B, the exon encoding the insertion in HNF-4α4 mRNA, revealed an additional T between nucleotides 219 and 220 in both alleles of five unrelated individuals (10 chromosomes) not present in the cDNA sequence[28], which causes a frameshift and generates a protein of 98 amino acids whose function is unknown. The sequences of the exons and adjacent introns have been deposited in the GenBank database (accession numbers, U72959–U72969).

Screening of the HNF-4α gene for mutations. The eleven exons and flanking introns were amplified using PCR and specific primers (the primer sequences are described in Supplementary Information). PCR conditions were denaturation at 94 °C for 5 min followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 30 s, with a final extension at 72 °C for 10 min. PCR products were purified using a Centricon-100 membrane (Amicon) and sequenced directly from both ends using an AmpliTaq Cycle Sequencing Kit and ABI Prism 377 DNA sequencer. The sequence of the Q268X mutation was confirmed by cloning the exon-7 PCR product into pGEM-T (Promega) and sequencing individual clones. The presence of the C → T mutation in codon 268 in members of the R-W pedigree and a group of unrelated non-diabetics was assessed by amplifying exon 7, digesting with *Bfa*I, and separation of the digested PCR product on a 3% agarose gel. The frequency of the T/I130 polymorphism was determined by amplification of exon 4 and digestion with *BstB*I, which does not cleave the Thr allele but cleaves the Ile allele into fragments of 184 and 87 bp.

Received 16 October; accepted 5 November 1996.

1. Fajans, S. S. *Diab./Metabol. Rev.* **5,** 579–606 (1989).
2. Froguel, P. *et al. N. Engl. J. Med.* **328,** 697–702 (1993).
3. Ledermann, H. M. *Lancet* **345,** 648 (1995).
4. Herman, W. H. *et al. Diabetes* **43,** 40–46 (1994) [erratum, *Diabetes* **43,** 1171 (1994)].
5. Byrne, M. M. *et al. Diabetes* **44,** 699–704 (1995).
6. Bell, G. I. *et al. Proc. Natl Acad. Sci. USA* **88,** 1484–1488 (1991).
7. Vaxillaire, M. *et al. Nature Genet.* **9,** 418–423 (1995).
8. Yamagata, K. *et al. Nature* **384,** 455–458 (1996).
9. Sladek, F. M., Zhong, W., Lai, E. & Darnell, J. E. Jr *Genes Dev.* **4,** 2353–2365 (1990).
10. Kuo, C. J. *et al. Nature* **355,** 457–461 (1992).
11. Mangelsdorf, D. J. *et al. Cell* **83,** 835–839 (1995).
12. Fajans, S. S., Bell, G. I., Bowden, D. W., Halter, J. B. & Polonsky, K. S. *Life Sci.* **55,** 413–422 (1994).
13. Irwin, M., Cox, N. & Kong, A. *Proc. Natl Acad. Sci. USA* **91,** 11684–11688 (1994).
14. Stoffel, M. *et al. Proc. Natl Acad. Sci. USA* **93,** 3939–3941 (1996).
15. Chartier, F. L., Bossu, J.-P., Laudet, V., Fruchart, J.-C. & Laine, B. *Gene* **147,** 269–272 (1994).
16. Xanthopoulos, K. G. *et al. Proc. Natl Acad. Sci. USA* **88,** 3807–3811 (1991).
17. Miquerol, L. *et al. J. Biol. Chem.* **269,** 8944–8951 (1994).
18. Chen, W. S. *et al. Genes Dev.* **8,** 2466–2477 (1994).
19. Emens, L. A., Landers, D. W. & Moss, L. G. *Proc. Natl Acad. Sci. USA* **89,** 7300–7304 (1992).
20. Zhang, X.-K., Salbert, G., Lee, M.-O. & Pfahl, M. *Mol. Cell. Biol.* **14,** 4311–4323 (1994).
21. Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H. & Moras, D. *Nature* **375,** 377–382 (1995).
22. Renaud, J.-P. *et al. Nature* **378,** 681–689 (1995).
23. Wagner, R. L. *et al. Nature* **378,** 690–697 (1995).
24. Byrne, M. M. *et al. Diabetes* **45,** 1503–1510 (1996).
25. Hanis, C. L. *et al. Nature Genet.* **13,** 161–166 (1996).
26. Iwasaki, N. *et al. J. Japan. Diab. Soc.* **39,** 409–416 (1996).
27. Polonsky, K. S., Sturis, J. & Bell, G. I. *N. Engl. J. Med.* **334,** 777–783 (1996).
28. Drewes, T., Senkel, S., Holewa, B. & Ryffel, G. U. *Mol. Cell. Biol.* **16,** 925–931 (1996).
29. Tavaviras, S., Monaghan, A. P., Schütz, G. & Kelsey, G. *Mech. Dev.* **48,** 67–79 (1994).

SUPPLEMENTARY INFORMATION is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as a paper copy from Mary Sheehan at the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. K.Y., H.F. and N.O. contributed equally to this work. This work was supported by the Howard Hughes Medical Institute, the US Public Health Service, Bristol-Meyers Squibb, the Juvenile Diabetes Foundation Internal, the Blum-Kovler Foundation, the Pew Charitable Trust and the Irma T. Hirschl Trust.

CORRESPONDENCE and requests for materials should be addressed to G.I.B. (e-mail: g-bell@uchicago.edu).

Role of learning in three-dimensional form perception

Pawan Sinha & Tomaso Poggio

E25-201, Center for Biological and Computational Learning, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 45 Carleton Street, Cambridge, Massachusetts 02142, USA

One of the most remarkable characteristics of the human visual system is its ability to perceive specific three-dimensional forms in single two-dimensional contour images. This has often been attributed to a few general purpose and possibly innately specified shape biases[1–6], such as those favouring symmetry and other structural regularities (Fig. 1). An alternative approach proposed by the early empiricists[7–10] and since tested[11] suggests that this ability may also be acquired from visual experience, with the three-dimensional percept being the manifestation of a learned association between specific two-dimensional projections and the correlated three-dimensional structures. These studies of shape learning have been considered inconclusive, however, because their results can potentially be accounted for as cognitive decisions that might have little to do with shape perception *per se*. Here we present an experimental system that enables objective verification of the role of learning in shape perception by rendering the learning to be perceptually manifest. We show that the human visual system can learn associations between arbitrarily paired two-dimensional pictures and (projectionally consistent) three-dimensional structures. These results implicate high-level recognition processes in the task of shape perception.

Our experiment comprised a training and a test phase (Fig. 2). For use as stimuli, we generated random two-dimensional (2-D) line-drawings and assigned them arbitrary three-dimensional (3-D) structures. During the training phase, subjects were shown one such object rocking through an angle of ±20 degrees about a frontoparallel horizontal axis passing through its centroid (we shall refer to this as the 'training' object). This was to allow the subjects to observe the correlation between the object's mean-angle projection and its associated 3-D structure through the kinetic-depth effect (KDE)[12,13]. The test phase commenced five seconds after the training session and was intended to assess subjects' shape learning. Subjects were shown either the training object rocking back and forth, or another object that had the same mean-angle projection but a completely different 3-D structure (the 'test' object). They were asked to indicate whether the objects looked rigid or non-rigid. We expected that if the subjects had indeed learned an association between the training object's mean-angle projection and its 3-D structure, then, when presented with a test object having the same mean-angle projection, they would perceptually impose on it the learned 3-D structure and the observed motion pattern would be mapped onto this learned structure. If in fact the test object had a different 3-D structure, the sequence would appear to depict a non-rigid deformation. We

also expected that switching the roles of the training and test objects across different populations of subjects would lead to reversals in the pattern of results. Furthermore, subjects who had not undergone the training session were not expected to perceive non-rigidity, because of the well known bias towards rigidity[12,13].

We divided our twelve subjects into three mutually exclusive equally sized groups. Members of the first group individually underwent a training phase and then a test session. Each subject was tested with ten different objects, each object being used only once. Members of the second population served as controls and did not undergo training sessions for any of the ten objects used. The procedure for the third population was identical to that for the first, apart from having the roles of the training and test objects reversed.

Figure 3 shows the results from the three populations. The results are plotted as the percentage of trials (over all objects) during which subjects perceived non-rigidity in the object depicted in the motion sequence. Members of population (1) show a clear effect of visual experience (Fig. 3*a*). Whereas the training object appeared to be non-rigid for an average (across all four subjects) of 7.5% of the presentations, the test object (rotating about the same axis) appeared to be non-rigid in ~40% of the presentations. Members of population (2), on the other hand, perceived both the training and the test objects as being rigid in most presentations (Fig. 3*b*). Members belonging to population (3) gave a reverse pattern of results (Fig. 3*c*).

The marked differences in the results of populations (1), (2) and (3) attest to the strong influence of visual experience on subsequent perception of 3-D form. We conclude that it is possible for the human visual system to learn associations between arbitrary 2-D projections and 3-D structures, and that this learning subsequently influences 3-D form perception. At least for the kinds of objects we consider here, mutual temporal correlation seems to be more important for the formation of such associations than any intrinsic structural characteristics of the 2-D and 3-D shapes. Supporting this interpretation, subjects reported recognizing the test object as being the one they had seen during the training session.

To characterize subjects' shape-learning better, we tested their performance under a few additional test conditions. As Fig. 3*d* shows, the learning is long-lasting. Two subjects from population (1) who were trained with 10-min sequences for each of five objects (five training sessions evenly spaced over 4 hours; every session comprised 2-min sequences for each of the five objects) showed significant effects of learning upon being tested a day later. The learning also seems to be object-specific in that the percept of non-rigidity declines with the addition of 2-D positional noise to the vertices of the original projection (Fig. 3*e*). However, transformations that preserve the shape of the training projection, such as image scaling, do not have a large effect on subjects' responses (Fig. 3*f*).

In the light of these results, an important issue that needs to be addressed is how the learned 3-D structure is represented by the visual system. The two main possibilities are: either the 3-D structure can be represented 'explicitly' as an object-centred CAD-like model[14], or it may be stored 'implicitly'—for instance, as a viewer-centred depth structure coupled with knowledge of how its projection transforms upon rotation about the training axis. To distinguish between these two hypotheses, we repeated our experiments with new test sequences that showed the test object rocking about an axis different from the training axis. For an explicit 3-D representation, we expected that the learning would manifest itself in the new test sequences and the percept of non-rigiditiy would be undiminished. An implicit 3-D representation, however, would predict a decrement in the percept because of a lack of transfer of learning across different rotation axes. As Fig. 3*g* shows, the percept of non-rigidity obtained with the new test sequences was significantly diminished relative to the original test sequence. This lends support to the idea of implicit 3-D representations, which is consistent with the suggestions of view-based representations for recognizing 3-D structures[15–18] and demonstrations of the ability of neurons in the infero-temporal cortex to code together temporally correlated frames in a motion sequence[19,20].

Figure 4 shows two illusions related to the ideas we have presented here. Figure 4*a* and *b* shows a situation in which a completely rigid wireframe object resembles a person, which when rocked around the vertical axis is perceived as a person walking (a

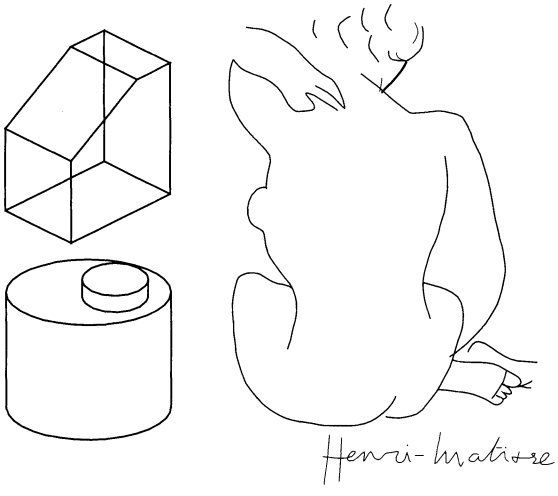

FIG. 1 These 2-D images convey a vivid sense of three-dimensionality. It is unclear whether the human visual system overcomes the underconstrainedness inherent in achieving this percept by the use of a few fixed and innately specified shape biases[1–6], or by learning associations between specific 2-D projections and the correlated 3-D structures[7–10,4–6]. Computational schemes that employ a small set of prespecified shape biases have been able to mimic human 3-D shape perception with images of relatively regular geometric objects, such as the two shown on the left. They prove ineffective, however, in more general domains that might contain images such as the one shown on the right.

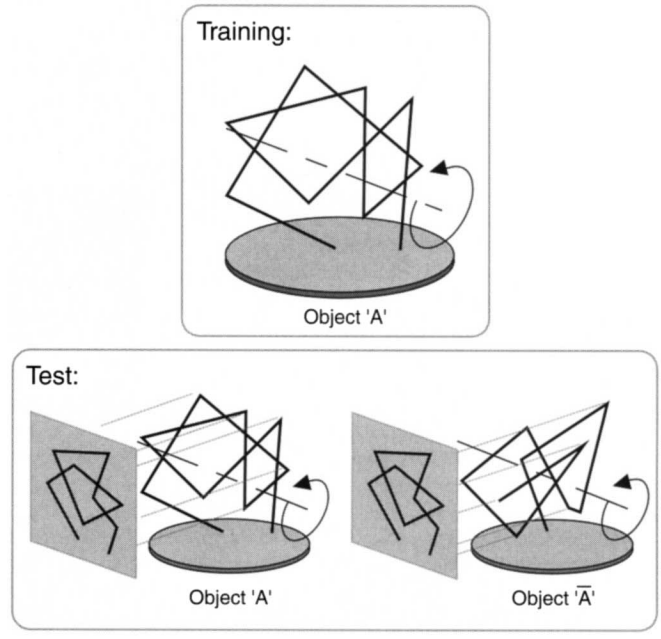

FIG. 2 Our experimental design. During the training phase, subjects saw a motion sequence of a rocking rigid 3-D object. In the subsequent test phase, subjects either saw motion sequences of the training object or of novel rigid objects, some of which had the same mean-angle projection as the training object. They were asked to report whether the objects looked rigid or non-rigid.

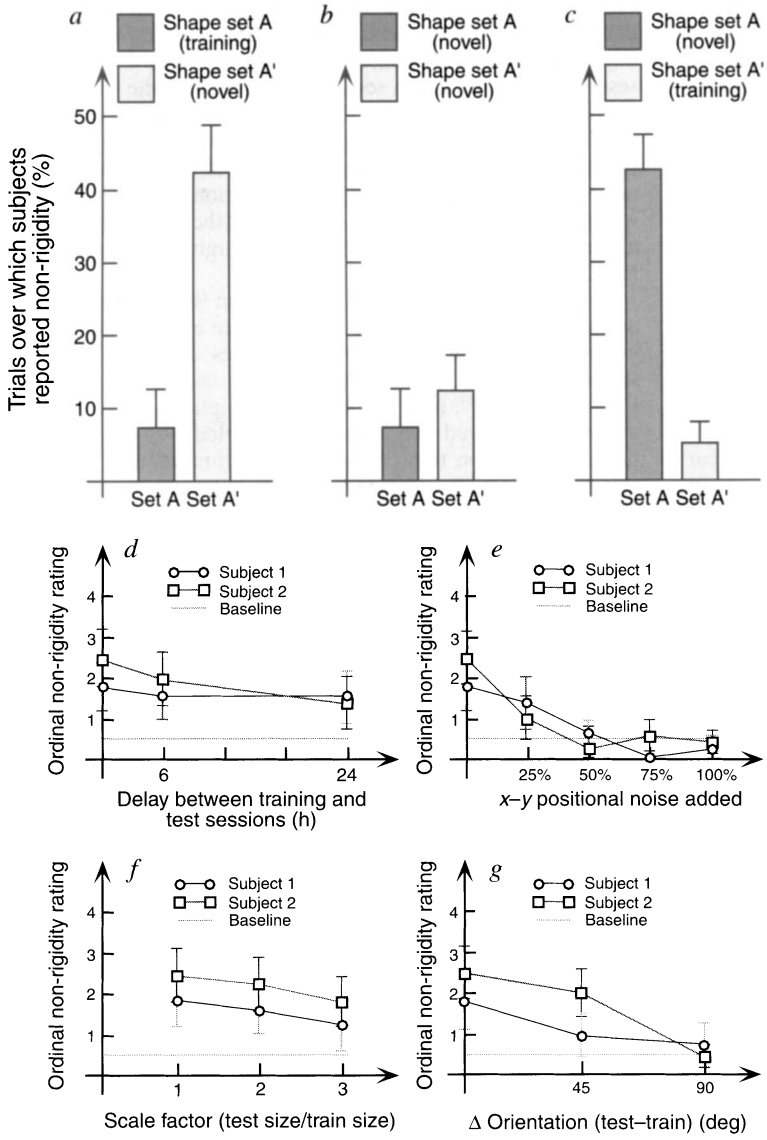

FIG. 3 Results from our three populations of subjects. Subjects were asked to indicate whether the objects shown in the motion sequences looked rigid or non-rigid. *a*, Results from population (1) on training and test objects. *b*, Results from population (2). Subjects in this group had not undergone any training sessions. *c*, Results from population (3). The training and test objects were switched for this population relative to population (1). For the experiments whose results are shown in *d*–*g*, subjects were instructed to rate the perceived non-rigidity on an ordinal scale with a score of zero corresponding to a very rigid percept, and '4' to a very non-rigid one. The grey horizontal line in all graphs is the baseline rating from a control group that did not undergo any training. The data are reported for one naive subject and one author. *d*, Persistence of shape-learning over time. *e*, The effect of adding 2-D positional noise to the mean projection of the test object. *f*, Effect of changes in image scale on the percept of non-rigidity. The training objects used here subtended, on an average, a visual angle of 2.5 deg. *g*, Decrement of the non-rigidity percept with changes in the rotation axis (angles indicate orientation differences in the image plane).

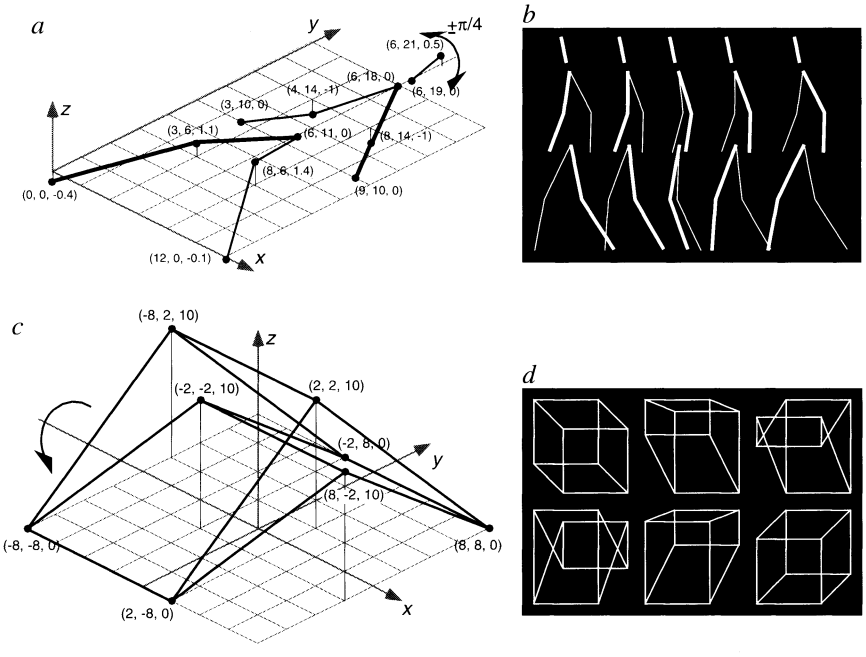

FIG. 4 Two illusions related to the ideas presented here. *a*, This rigid structure, when rocked back and forth through an angle of 90 deg about the vertical axis, is perceived as a human walking (a non-rigid interpretation). *b*, Frames from the motion sequence. This mismatch between reality and perception is probably due to the recognition of the figure as a human, which leads the visual system to interpret the motion patterns in terms of how the human structure is expected to change. This expectation is likely to have been learned through visual experience. The same vertex set, when not recognized as a human (say, when the vertices are connected up randomly) appears rigid upon rotation. *c*, An object that projects to a cube (from a specific viewpoint) but has a very different 3-D structure. *d*, A motion sequence of this object, that includes the cube-like projection, appears to depict a highly non-rigid object. With the vertices connected randomly, the percept of rigidity is restored. Just as for *a*, we suggest that the perceived non-rigidity here is likely to be due to the mismatch between the observer's expectations and the presented sequence. It is not clear whether the shape expectation for a cube (and other simple geometric objects) is learned during an observer's lifetime, or is innate. (Copies of these sequences are available from the authors.)

non-rigid interpretation). In this case, the association learned in the course of our daily experience overrides the general bias towards a rigid interpretation. Figure 4*c* shows a non-cuboidal object whose mean 2-D projection is the same as a cube's. A KDE sequence of this object rotating (Fig. 4*d*) looks highly non-rigid. However, if we connect the vertices of this object in a random fashion so that it no longer looks like a cube, thus weakening the shape expectations, its structural rigidity becomes evident.

The idea that 3-D shape perception might be mediated, at least in part, by object-specific learning has a direct corollary: 3-D shape perception may sometimes follow recognition rather than precede it (as the visual system has to be able to recognize the 2-D projection to access its associated 3-D structure). This idea (also suggested by Cavanagh[21]) runs contrary to the traditionally assumed hierarchy of visual processing, according to which recognition depends upon the results of the 3-D-shape recovery process[22]. It is natural to ask whether the learning reported here could generalize to novel objects of the same object class, as suggested by a computational scheme based on linear combinations of prototypes (ref. 23 and included references). Preliminary experiments to be reported elsewhere suggest that this is indeed the case and that the object-specific learning we describe can generalize to similar objects. □

Methods

Subjects were trained on a 60-s sequence showing a randomly constructed rigid 3-D object rocking back and forth through an angle of 40 degrees with an angular speed of 30 deg per second. The test objects rocked back and forth through an angle of 20 deg, with an angular speed of 30 deg per second for 6 s. All sequences were presented on a Silicon Graphics workstation. The stimuli had the form of a thin wire randomly bent at seven arbitrary places along its length, displayed using anti-aliased white segments on a black background. There were no occlusion or shading cues. On average, they subtended a visual angle of 6 deg from a viewing distance of 80 cm. The objects were selected to yield a rigid percept to a group of three naive observers and also to ensure that they did not present views with accidental alignments of the constituent segments. All subjects except one (author P.S.) were naive as to the purpose of the experiment.

Received 28 June; accepted 10 October 1996.

1. Koffka, K. *Principles of Gestalt Psychology* (Harcourt Brace, New York, 1935).
2. Kohler, W. *Physical Gestalten* (reprinted in *A Sourcebook of Gestalt Psychology*) (ed. Ellis, W. D.) (Humanities, New York, 1920).
3. Hochberg, J. & McAlister, E. *Exp. Psychol.* **46,** 361–364 (1953).
4. Attneave, F. & Frost, R. *Perception Psych.* **6,** 391–396 (1969).
5. Kanade, T. *Artificial Intelligence* **17** (1–3), 409–460 (1981).
6. Fischler, M. A. & Leclerc, Y. G. *Proc. DARPA Image Understanding Workshop* (Morgan Kaufmann, San Francisco, CA, 1992).
7. Locke, J. *An Essay Concerning Human Understanding* (1690) (in *The English Philosophers from Bacon to Mill*) (ed. Burtt, E. A.) (Modern Library, New York, 1939).
8. de Condillac, E. B. *Treatise on the Sensations* (1754). In: Philosophical writings of Etienne Bonnot, abbe de Condillac. Translated by Franklin Philip and Harlan Lane. (Lawrence Erlbaum Associates, Hillsdale, NJ, 1982).
9. Hume, D. *A Treatise of Human Nature* (Everyman's, London, 1738/1956).
10. von Helmholtz, H. *Helmholtz's Physiological Optics* (1866/1911). Translated for mthe third edition (1909–1911) by J. P. Southwell (ed.), Rochester, New York: Optical Society of America.
11. Wallach, H., O'Connell, D. N. & Neisser, U. *J. Exp. Psychol.* **45,** 360–368 (1953).
12. Wallach, H. & O'Connell, D. N. *J. Exp. Psychol.* **45,** 205–217 (1953).
13. Ullman, S. *The Interpretation of Visual Motion* (MIT Press, Cambridge, MA, 1979).
14. Marr, D. & Nishihara, H. K. *Proc. R. Soc. Lond. B* **200,** 269–294 (1978).
15. Rock, I. & Di Vita, J. *Cognitive Psychology* **19,** 280–293 (1987).
16. Poggio, T. & Edelman, S. *Nature* **343,** 263–266 (1990).
17. Tarr, M. J. & Pinker, S. *Psychol. Sci.* **2,** 207–209 (1990).
18. Bülthoff, H. H. & Edelman, S. *Proc. Natl Acad. Sci. USA* **89,** 60–64 (1992).
19. Miyashita, Y. *Annu. Rev. Neurosci.* **16,** 245–263 (1993).
20. Stryker, M. *Nature* **354,** 108–109 (1991).
21. Cavanagh, P. in *Representations of Vision: Trends and Tacit Assumptions in Vision Research* (ed. Gorea, A.) (Cambridge University Press, Cambridge, 1991).
22. Marr, D. *Vision* (Freeman, San Francisco, 1982).
23. Beymer, D. & Poggio, T. *Science* **272,** 1905–1909 (1996).
24. Morgan, M. J. *Molyneux's Question: Vision, Touch and the Philosophy of Perception* (Cambridge University Press, Cambridge, 1977).
25. Ames, A. *Psychol. Monogr.* **65,** 1–32 (1951).
26. Gregory, R. L. *Perceptions as Hypotheses* Ch. 9 (ed. Brown, S. C.) (Macmillan, London, 1974).

ACKNOWLEDGEMENTS. We thank H. Bülthoff, P. Cavanagh, M. Fahle, R. Held, A. Parker, M. Potter, S. Ullman and T. Vetter for reading the manuscript and for insightful comments, and A. Hurlbert and M. Morgan for constructive suggestions. The walking figure illusion described in Fig. 4 was created in collaboration with P. Lipson. This research was supported by grants from the Office of Naval Research, NSF and DARPA. P.S. is supported by a McDonnell-Pew postdoctoral fellowship; T.P. is supported by the Uncas and Helen Whitaker Chair at the Whitaker College, MIT.

CORRESPONDENCE and requests for materials should be addressed to P.S. (e-mail: sinha@ai.mit.edu).

Estimation of self-motion by optic flow processing in single visual interneurons

Holger G. Krapp & Roland Hengstenberg

Max-Planck-Institut für Biologische Kybernetik, Spemannstrasse 38, D-72076 Tübingen, Germany

HUMANS, animals and some mobile robots use visual motion cues for object detection and navigation in structured surroundings[1–4]. Motion is commonly sensed by large arrays of small field movement detectors, each preferring motion in a particular direction[5,6]. Self-motion generates distinct 'optic flow fields' in the eyes that depend on the type and direction of the momentary locomotion (rotation, translation)[7]. To investigate how the optic flow is processed at the neuronal level, we recorded intracellularly from identified interneurons in the third visual neuropile of the blowfly[8]. The distribution of local motion tuning over their huge receptive fields was mapped in detail. The global structure of the resulting 'motion response fields' is remarkably similar to optic flow fields. Thus, the organization of the receptive fields of the so-called VS neurons[9,10] strongly suggests that each of these neurons specifically extracts the rotatory component of the optic flow around a particular horizontal axis. Other neurons are probably adapted to extract translatory flow components. This study shows how complex visual discrimination can be achieved by task-oriented preprocessing in single neurons.

For several decades, coherent retinal image-shifts induced during locomotion have been considered a rich source of information about the momentary self-motion[7]. These retinal image-shifts, known as optic flow fields, can be described as vector fields where each vector indicates the direction and velocity of the local image-shift[11,12]. The structure of an optic flow field depends on the instantaneous self-motion[12], which can be described in terms of its rotatory and translatory components. Electrophysiological[13,14] and behavioural experiments[15] in different animal species and psychophysical studies in humans[3,16] strongly suggest that optic flow is indeed exploited for the estimation of self-motion.

Globally, translatory and rotatory optic flow fields can be distinguished easily (compare Fig. 1*a* and *b*). However, in the first processing stages of the visual system, motion is analysed locally by elementary motion detectors (EMDs)[5,6,17]. Different kinds of self-motion may induce the same excitation in an EMD because the local flow vectors are quite similar or even equal at the respective position (see framed vectors on the right in Fig. 1*a* and *b*). Consequently, the two different self-motions cannot be distinguished at the level of the local EMD.

One simple way of overcoming these ambiguities is to integrate spatially the signals of those EMDs whose preferred directions correspond with the direction of the local velocity vectors of a particular optic flow field. In this case, the integrating neuron would respond best when the animal performs the corresponding self-motion in space. If this strategy were realized in the visual system, the integrating neurons should satisfy the following requirements: (1) the neurons should have extended receptive fields—because the ability to distinguish between different flow fields increases with increasing area of integration[12]; (2) the neurons' responses have to be motion-sensitive and direction-selective; (3) the local preferred direction of the neuron should depend in a characteristic way on the position of the local motion stimuli in the visual field.

The third visual neuropile (lobula plate) of the blowfly contains about 60 so-called tangential neurons that are individually identifiable from animal to animal[8,9]. Most of these neurons receive retinotopic input[18], on their extended dendritic arborizations, from many local EMDs. Microsurgical and neurogenetic deletions