Artefact or network evolution?

SIR — Enquist and Arak[1,2] and Johnstone[3] have chosen an inappropriate methodology to demonstrate principles of human perception of two-dimensional patterns in using neural nets with multi-dimensional input layers. When their input stimuli are displayed in two-dimensional arrays (a, c in the figure), the human eye sees symmetries and other spatial properties, but their artificial networks did not in fact use such spatial information in determining the correct response.

The input layers were actually n-dimensional (36-, 147- and 20-dimensional, respectively); their display as two-dimensional "retinal arrays" was therefore entirely arbitrary[4] and had no relevance to network performance. The 6×6 input array in ref. 1 could equivalently be shown as a 4×9 array, and the 4×5 input array in ref. 3 could be displayed as a 5×4 array. If such alternative displays are chosen, the pattern labelled as a "conspecific bird" would appear to biological visual systems as remarkably un-bird-like (b in the figure) and the pattern labelled as the "most symmetrical tail" would appear remarkably asymmetrical (d). In other words, the networks did not learn anything about two-dimensional patterns — conspecific, symmetrical or otherwise — but they did learn the frequency with which individual input pixels indicated the correct output response. The simulation results in all three studies can therefore be explained entirely on the basis of the correlation coefficients between input and output units[5].

In Johnstone's study[3], the φ correlation coefficients[6] between input and output units were strongly positive for all pixels in the central two columns in c. Whenever any such pixel was activated in trained networks, it was an indication that the pat-

a, "Conspecific male bird with long tail" when displayed in a 6×6 array[1]; b, the same "bird" when displayed in a 4×9 array; c, the "most symmetrical tail" when displayed in a 4×5 array[3]; d, the same "tail" displayed in a 5×4 array.

tern was a "tail", regardless of the two-dimensional configuration of the tail. For example, activation of pixel 2 indicates the presence of a "tail" in all five training patterns and never (or very infrequently, depending on the randomization technique) indicates a "non-tail". This is why the most symmetrical tail (c) was recognized as a tail even when it was not included in the training set: all its pixels individually had strong correlations with the desired "tail" output.

Again, in Enquist and Arak's study[1], the pattern of φ coefficients rather than ideas about the perception of tail length explain the network results, whereas in their more recent simulation[2] point-biserial correlation coefficients[6] between input and output units suffice to explain the evolution of symmetrical patterns. There, the nets were designed to select stimuli with strong primary colours, so that the rotation and reflection of input stimuli tended to produce strong synaptic weighting from input units located at positions 90, 180 and 270° from such inputs. The strong synaptic weighting in turn favoured selection of signals with the appropriate colour at corresponding signal positions, but these two-dimensional patterns, which biological visual systems see as symmetries, are, to such neural nets, nothing more than strong inputs in the 147-dimensional input space. Their positions in the input layer are irrelevant and their symmetry when displayed as a specific two-dimensional pattern plays no part in network performance[5].

The problem in all three simulations is that neural networks are quick to exploit the correlational structure of the input–output vectors, whereas spatial forms rather than absolute retinal positions are the essence of perception by biological systems. As a consequence, although all the evolutionary arguments made in these three studies[1–3] may well be valid for other reasons, conclusions about the perception of two-dimensional patterns (objects, symmetries, tail-length, and so on) cannot be drawn on the basis of neural networks that have n-dimensional input layers.

Norman D. Cook
Faculty of Informatics, Kansai University, Takatsuki, Osaka 569, Japan

JOHNSTONE REPLIES — Cook argues that the networks in my simulation[3] did not learn anything about symmetry; rather, they merely learned the frequency with which individual pixels indicated the correct or incorrect output. Consequently, he suggests, the most symmetrical tail pattern elicited a strong response, not because of its symmetry (which played no role in network performance), but simply because it comprised pixels that were common to the set of tail patterns, and thus had a strong positive correlation with the desired 'tail' output.

I entirely agree with this argument; in fact, I made the same point in my paper: trained networks preferred symmetrical tail patterns because these were close to the average of the training set, not because they were symmetrical per se, a point reinforced by the observation that random symmetrical images differing substantially from the training set average were not preferred.

The reason the symmetrical tail pattern was closest to the training set average (that is, comprised those pixels common to the largest proportion of training patterns) was that the asymmetric tail patterns in the training set formed mirror-symmetric pairs. Because paired ornaments in nature, such as swallows' tails or earwigs' forceps, also show symmetrical variation (FA being defined as deviations from symmetry that are random with respect to side), symmetrical males are also closer to the population average display. They, too, may be preferred simply for this reason, as a by-product of selection for mate recognition. This point is of some interest to evolutionary biologists, because it implies that preferences for symmetrical males can be explained without necessarily invoking a link between male symmetry and quality.

R. A. Johnstone
Department of Zoology,
University of Cambridge,
Downing Street, Cambridge CB2 3EJ, UK

ENQUIST AND ARAK REPLY — Cook suggests first that we have made the mistake of assuming that two-dimensional symmetry is information that is used by our nets[2]; and second, he raises a more general question of whether the particular neural nets we used really 'see' patterns in the same way that humans see them.

On the first point we agree with Cook that our neural nets did not code information about the geometrical structure of input patterns. This would be a serious criticism of our paper had we claimed that the networks could accomplish tasks of spatial perception, or detect symmetry per se in patterns, which clearly they cannot.

We started our simulations with an arbitrary pattern of input, and a pseudo-retina consisting of a number of cells arbitrarily arranged on a two-dimensional surface. As long as we assume the positions of input cells are fixed in space (as they are in a real retina), the conclusion that follows from our simulations is inevitable: in the process of selecting nets
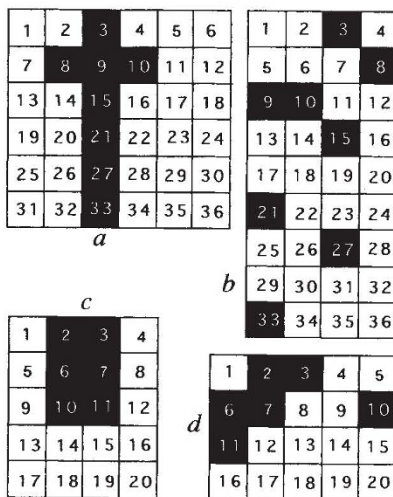
1. Enquist, M. & Arak, A. *Nature* **361**, 446–448 (1993).
2. Enquist, M. & Arak, A. *Nature* **372**, 169–172 (1994).
3. Johnstone, R.A. *Nature* **372**, 172–175 (1994).
4. Scalettar, R. & Zee, A. *Biol. Cyb.* **58**, 193–201 (1988).
5. Cook, N.D., Früh, H. & Landis, T. *J. exp. Psychol.: Hum. Perc. Perf.* (in the press).
6. Kurtz, A.K. & Mayo, S.T. *Statistical Methods in Education and Psychology* 328–356 (Springer, New York, 1979).