Commentary/Strong & Whitehead: Neural networks

The Y cells project both to the superior colliculus and, via the Y-magnocellular pathway, to the visual cortex, where they provide input to the motion-detection system. This pathway in turn provides information on the spatial locations of potential targets to the oculomotor system via both the internal corticothalamic tract from the extrastriate cortex and the projection from the frontal eye fields (reviewed by Gouras 1985). Thus the Y-magnocellular pathway also encodes spatial affordances in the sense required by S & W's model. As signals flowing along this pathway are available for integration into the pattern analysis system, affordances encoded in these signals could serve as tags.

The various stages of the Y-magnocellular pathway, not the burst cells of the superior colliculus, are thus the more appropriate loci for the representation of spatial affordances in the model. This interpretation is more consistent with S & W's simulation, in which affordance detection is carried out by simulated cortical minicolumns (sect. 3.4), than is the collicular interpretation of affordances that motivated the model.

Because Y cells respond transiently, the generation of a representation of the spatial array of stimuli stable enough to be useful as a tag for the relocation of particular purported objects requires the integration of correlated inputs over time (i.e., over multiple eye movements). Thus the process of tag generation itself requires (transient) tagging of locations as locations of stimuli based on input correlations over time. Such integration by correlation detection is presumably made possible by the large receptive fields of the Y cells, and the even larger receptive fields of many of the cells of the motion pathway. This process may not be completed until the level of the parietal cortex. This suggests, at any rate, that the construction of spatial tags is a temporally continuous process that modulates, and is modulated by, the identification of objects, and that is executed by the multiple motion areas of extrastriate cortex. If this is the case, the "affordances" that are detected, and that serve as targets for saccades, are not likely to be points in static layouts, but rather positions over time of partially interpreted objects. Such affordances may better serve the role of targets of appropriate action that Gibson's theory (1979) assigns them.

ACKNOWLEDGMENT

This work was partially supported by NASA Innovative Research Program grant NAGW-1592 to the author and J. Barnden.

More packaging needed before tags are added

John Findlay and Robert Kentridge
Department of Psychology, University of Durham, Durham DH1 3LE, England
Electronic mail: j.m.findlay@durham.ac.uk

Strong & Whitehead (S & W) address the problem of relating objects and their features to a spatial framework. We believe that they are right to see the solution to this problem as a fundamental requirement for a theory of vision. Their approach to the problem involves an associative network in which the spatial tags are achieved through reference to the affordance for saccadic eye movements. This seems a novel and promising approach. However, the specific model presented in support of the argument has a number of limitations and in some of its key features is also seriously at variance with empirical data. In the S & W model, "objects" become defined as such through the synchronous firing of minicolumns associated with their constituent features during learning. This presents no problem in S & W's visual environment. They believe that the pattern is unique and so can serve as a spatial tag to distinguish the location of the fixation.

What spatial behaviour would the model support? S & W point out that the same location could be identified on refixation, accepting for the moment that the affordance pattern is unique. However, this property alone would not seem to suffice for other desirable spatial properties. It seems important, for example, to have a procedure to refixate a previously fixated location from any other fixation location. As suggested in the free-running stage of the simulation, the associative network could, from a point in the feature space, recover a fixed set of spatial coordinates (or motor error signals). However, the spatial coordinates in the S & W model are the affordances present in the field when the fovea is centred, not ones which automatically take the current fixation position into account (to do this might be possible using a tensor transformation along the lines of Pelliövali & Linjas 1985).

It is clear that the visual signal leading to saccadic eye movements has some important properties. These have been investigated by physiological and psychophysical studies of saccadic eye movements in relation to targets which make a sudden appearance in the visual field; it seems possible that they also apply to the signals arising when saccades are made in a stationary visual environment. As S & W have noted, in the case of a single target, the saccadic premotor signal encodes the target spatial position independently of the target features. This has led to their model of spatial encoding, which takes the form of a parameter net. As defined in the target article (and rather differently in the preprint of Ballard 1987), a parameter net is a winner-take-all network of minicolumns within a spatial array. This representation is at variance with the empirical data. Both physiological and psychophysical studies suggest that the spatial signal for saccadic eye movements is encoded in a distributed manner in the premotor centres and furthermore uses a value-averaging system rather than a winner-take-all system. A crucial finding is the "global effect." When two targets are simultaneously presented in the visual field, the resultant saccade is directed to an intermediate position which may be described as the "centre of gravity" of the targets (Coren & Hoenic 1972; Findlay 1982; Ottes et al. 1984). The most plausible interpretation of this result is that the visual information undergoes extensive spatial integration (Deubel et al. 1984, Findlay 1987). Recent work has shown that the properties which are integrated are both local properties of the stimulus and properties such as contours and texture boundaries, which are likely to be important in early visual figure-ground processing (Deubel et al. 1988). Thus the "picture" sent from the visual system to the saccadic system does not consist of a discrete set of saccade targets but is a highly convoluted representation of the visual input.

S & W appear to be committed to synchronous firing as a way of defining the emergence of a cell assembly. This agrees with experimental evidence for the existence of synchronised assemblies such as that presented in Johannesma et al. (1986). S & W's model assumes that the minicolumn is the unit of selection in assembly formation (i.e., an assembly is defined in terms of a set of simultaneously active minicolumns rather than a set of simultaneously active neurons). This is potentially a problem. The emergence of assemblies of synchronously firing neurons from background activity depends on the fact that neurons are an order of magnitude more sensitive to synchronous than to asynchronous presynaptic excitation (Abeles 1982, p. 68). There...
is no reason to suppose that minicolumns as units are subject to
the same dissociation of responses between synchronous and
asynchronous simulation. S & W's thresholding mechanism
operating through gap junction synapses within a minicolumn
appears to be intended to provide such sensitivity to syn-
chronous stimulation. It is in fact likely that gap-junction syn-
apses play a role in intercolumnar inhibition (Colonnier 1981)
rather than in intracolumnar excitation.

Simulations such as those of Gerstein et al. (1982) and analys-
es such as those presented in Abeles (1982) indicate that gap-
junction synapses are unnecessary for the formation of syn-
chronised assemblies. Furthermore, one of the strengths of
Abeles's (1982) synfire chains is that they are not sensitive to the
unreliability of individual neurons cited by S & W when they
suggest minicolumns as units of selections. It is possible, how-
ever, that minicolumns are still fundamental processing units in
that they respond selectively to characteristics of the patterning
of intercolumnar activity. Synfire chains formed during coinci-
dent activity of a number of minicolumns would be the result of
neuron-neuron interactions, but would be determined by the
(nonadaptable) sensitivities of the minicolumns of which they were
members.

The bimaceral retina at a glance

C. L. Hardin

Department of Philosophy, Syracuse University, Syracuse, NY 13244

As far as I can judge, Strong & Whitehead's (S & W's) post "reti-
nal" neural network modeling is biologically well motivated and a
real step forward in addressing the "tag" problem. But it might
be well to speak instead of "tag problems," for there seem to be
at least two: how feature complexes are identified across eye
movements, and how features and feature complexes are con-
ected to locations in the visual field. Although these two
problems have common characteristics (a solution to the one is
likely to be helpful in providing a solution to the other), there
are important differences between them, and it is worthwhile to
keep these differences in mind. In particular, an examination of
the reports of the experiments on attention and feature integra-
tion performed by Treisman and her associates indicates that (1)
there is a significant attachment of features to visual field
locations, one that occurs in periods too brief for sacades to be a
significant factor, and (2) even in complex serial searches requir-
ing up to 3 seconds, eye movements probably play a subordinate
role. Let us consider these points in turn.

The minimum interval between sacades is about 225 milli-
seconds. Treisman tells us, "In one experiment my colleagues
and I flashed three colored letters, say a blue X, a green T and a
red O, for a brief period (200 milliseconds) and diverted our
subjects' attention by asking them to report first a digit shown at
each side of the display and only then the colored letters. In
about one trial in three the subjects reported the wrong com-
binations - perhaps a red X, a green O or a blue T" (Treisman
1986, p. 117). Her purpose in saying this was of course to em-
phasize just how frequently the misperception of feature
location occurs in brief episodes of saccade. But what we should
also notice is that the assignment of features to locations in this
experiment was made correctly by the subjects in two-thirds of
the trials. Some sort of mechanism is making a fairly effective
assignment over a brief interval of time, an interval too short for
eye movement to play any role at all. This result is typical of
other brief-exposure experiments: one should therefore expect
that the performance by sacades in searches of 400 or 500
milliseconds to be minor.

But what of the searches that require much longer times - up
to 3 seconds? Treisman and Gelade (1980) commented on
one such experimental search:

Commentary/Strong & Whitelhead: Neural networks

Subjects certainly changed fixation and scanned the display with their
eyes, so that different areas of the display received foveal processing
successively. In this sense processing was serial. However, serial eye
fixations do not imply serial decisions about each item, one at a time,
and we believe the two patterns have different theoretical impli-
cations which are worth distinguishing. Serial fixations will be made
when the discriminations require foveal activity, either because they
are below threshold with peripheral vision or because of the specific
form of lateral interference which increases toward the periphery.
However, with each successive fixation it is at least logically possible
that the whole display receives parallel processing, the foveal areas
receiving the most detailed sensory information, but all or many stimulii
being checked simultaneously. Since density increased with
number of items in the present experiment, more stimuli would on
average have been within foveal vision for each fixation with the
larger display sizes, allowing the number that could be accurately
processed in parallel to increase with display size. . . . A search task
that requires the identification of conjunctions depends on a more
central scan with focused attention, which deals serially with each
item rather than with each spatial area foveally fixated. In this
case the difficulty should be restricted to conditions in which more
than one item is presented, allowing the possibility of feature interchanges
or "illusory conjunctions." But the search task is not as likely as the
limits set by acuity. Only the number of items should affect search
times, and not their density or spatial distributions. (Treisman &
Gelade 1980, pp. 112–13)

One might also ask whether the simplifications that S & W
introduce into their retinal model are too drastic. They assign to
their model fovea the single function of "nonspatial" feature
discrimination and to their model peripheral retina the single
function of providing "spatial information." Readily conceding that "this
differential ability is probably a graded function in human visual
processing" they add that "our model will not take this con-

nuity into account." However, they seem to insist that there is
such a division of function, albeit graded, between fovea and
periphery in real eyes.

The evidence that they provide for this division is not compel-
lng. They quite correctly tell us that there is "evidence for such
a separation of location information from nonspatial features in the
cortical pathways taken by the processing of the two kinds of
information." But of course this does not entail a corresponding
division of function between retinal regions. Neither does the
distinction between magnocellular and parvocellular systems to
which they draw our attention imply a division between areas of
retinal regions engaged in spatial and areas engaged in nonspatial
information processing. Although it is true that retinal cells project-
ing to magnocellular layers of the lateral geniculate nucleus
seem absent in the fovea, where parvo-projecting cells are
abundant, it is also true that in the periphery where magno-
projecting cells are found, parvo-projecting cells still out
number them by about five to one (Lennie & D'Zmura 1988, p.
366). S & W tell us that "information from the magnocentric
system alone is used for motion and spatial analysis." But the spatial
analysis in which the magno system is implicated seems to be
largely concerned with stereoscopic depth information, which
does not play a part in the S & W model. The parvo system, on
the other hand, does seem to be very much involved in spatial
analysis, particularly the resolution of spatial detail.

Resolution of spatial detail is just what is required for the
detection of many Treisman-type "features" such as the tail on
the letter "Q." (Considerations of this sort render suspect any
simple dichotomy between "feature analysis" and "spatial analy-
sis." ) The fovea is better suited for accurate spatial discrimina-
tions than regions of extrafoveal retina iffor no other reason than
that its receptors are tightly packed and have less tendency to
pool outputs with neighboring receptors than do their cousins in
the periphery. In the (approximately) two degrees of foveal
retina a large portion of spatial resolving and locating has its
start. How odd it is, then, for S & W to assign to it the sole task of
detecting nonspatial features! The prototypical "nonspatial

behavioral and brain sciences (1989) 12:3

405

405