Behavioral and Brain Science, vol 9, March 1986

[COMMENTARY ON DANA BALLARD'S ARTICLE:
"CORTICAL CONNECTIONS AND PARALLEL PROCESSING"]

****

Two Tests for the Value-Unit Model: Multicell Recordings and Pointers

****

David Mumford
Department of Mathematics
Harvard University
Cambridge, MA, 02138

I have two comments on the "value-unit" model presented in Ballard's article. The first concerns the implications of the data obtained from single-cell recordings in supporting the value-unit hypothesis. The second concerns the adequacy of value-units to deal with the situation in which a higher level concept must play two roles simultaneously in interpreting a stimulus.

More precisely, the first point is that when the only available data on the function of some area is from single cell recordings, this already pushes one strongly towards a value-unit interpretation. It is very difficult without using multicell recordings to compare the value-unit model with most of its alternatives. If one can only measure a single variable of the internal state of a complicated machine, unless one has a very good theory of what is happening inside this machine, the only case in which the time behavior of this variable is interpretable at all is going to be the case in which the variable is directly correlated to some clear feature of the input of the machine, i.e. when the variable can be interpreted as a value-unit. The possibility of adopting an over-simplified value-unit interpretation is further increased by the standard practise of averaging over many responses.

What are the alternative possibilities? Ballard mentions value-encoding. Another possibility is that the neurons in a local circuit are doing something complicated together which is hard to interpret, but that the output signal sent by some pyramidal cell a relatively long distance has a clear "value interpretation". Think for a moment quite generally how a massively parallel but only locally highly connected network might function. The hypothesis that individual units represent interpretable values is quite special; but the hypothesis that the relatively sparse long-distance interconnections send messages about interpretable data is hard to avoid. To go back to the data, an averaged single cell recording may often wipe out all structure except one particular phase of the response, which may well be roughly correlated to the output of the local circuit. In contrast, a multicell recording might be able to detect something of the time sequence of operation of this local circuit. Yet another alternative to neuronal value-units is that the computed data of a local circuit is represented by the simultaneous state of several neurons (or the output on several axons), i.e. the value-units are clusters of neurons. Clearly a multicell recording (cf. Gerstein et al (1983)) is needed to detect this.

I want to make this point more specific by reference to particular visual regions in the Macaque monkey. The hierarchy of simple/complex/hypercomplex cells is the best known set of low-level value-units. The "hand" and "face" cells in area IT discovered by Gross, Rocha-Miranda & Bender (1972) is a striking example of high-level units. But there is a big gap between these two of missing value interpretations for neurons in V3,V4,TEO which are apparently concerned with shape but for which "best" stimuli have not been found.

I'd like to conjecture that value interpretations of single neurons here don't exist, but that one needs to look at several at once to see what their response means. More specifically one could prepare a set of several hundred black and white shape stimuli such as outlines of leaves, animals, letters, polygons, blobs fo various types; record several nearby cells simultaneously responding to each shape; then perform a multi-dimensional cluster analysis on the responses to see if the stimuli are sorted in a coherent way by the multi-dimensional response of the group of cells being recorded from.

The second comment is related to the problem raised by Ballard of short-term memory. The basic problem is that higher-level value-units are trying to play the two roles of i) instantiating general categories and ii) being tokens for individual exemplars, both at once. To illustrate what I mean, suppose one is dealing with a higher-level unit whose value might be "triangle" or "hand". Consider images of the following kinds:

i) Images with 2 triangles or 2 hands, each of which must be individually categorized,

ii) Images with a shape which is a triangle or hand except for some missing element. Then the triangle or hand must be referred to with a footnote about what is missing.

iii) Images with a triangle or hand which is "part-of", (e.g "the-top-of") some bigger shape such as another triangle or hand.

In all these cases, specially elaborated triangles or hands must be described by setting up associations. This would seem to demand a system of pointers or a mechanism for referring to the high-level unit, making it possible to link it simultaneously into an interpretation of an image in many ways at once. Several ways to do this have indeed been proposed e.g. by Feldman (1981), but they seem to go beyond the pure idea of value-units.

To take an example, how would Gross's hand detectors cells respond when there are several hands in the visual field? Experimentally, one could arrange for the different hands to disappear and reappear and see how the cell reacts. One possibility would be that any particular cell choses one hand and ignores the others. This would tie up nicely with the pyramid-linking algorithms e.g. of Hong & Rosenfeld (1984), in which higher units are dynamically linked to parts of the visual field depending on content. It is similar to the solution in Figure 20 of the cross-talk problem. But if one believes in pure value-units it means that you must have a whole set of hand value-units available. The first one is ready to represent the presense of the first hand detected in the visual field, the second unit stands ready to respond to the second hand, etc.! This sounds somewhat implausible.

To summarize, it is hard for me to imagine how higher-level categories can be instantiated by a relatively small set of pure value-units, (even when coupled with a small set of auxiliary variables in "parameter subspaces"), and still be able to fulfill their many cognitive functions, linking in manifold ways with other structures. It seems necessary to introduce general purpose units or pointer substitutes to instantiate the associations themselves. How this is done in the brain seems to me to be one of the fundamental problems of neuroscience.

****

REFERENCES FOR THIS COMMENTARY:

Feldman,J. (1981), A connectionist model of visual memory, In Parallel models of associative memory, ed. G.E.Hinton & J.A.Anderson, pp.49-81, L.Erlbaum.

Gerstein,G.L., Bloom,M.J., Espinosa,I.E., Evanczuk,S., & Turner,M.R. (1983), Design of a laboratory for multineuron studies, IEEE Trans. on Systems, Man and Cybernetics, 13: 668-676.

Gross,C.G., Rocha-Miranda,C.E.,& Bender,D.B.(1972), Visual properties of neurons in inferotemporal cortex of the Macaque, J. of Neuro-physiology, 35: 96-111.

Hong,T.H. & Rosenfeld,A. (1984), Compact region extraction using pixel linking in a pyramid, IEEE Trans. on pattern analysis and machine intelligence, 6:

****