

Concept neurons: A proposed developmental study

WAYNE A. WICKELGREN
University of Oregon, Eugene, Oregon 97403

A neurophysiological experiment is proposed to determine the existence of visual object concept neurons. The basic technique is to deprive newborn animals of all visual experience with the exception of a small number of objects exposed one at a time in a Ganzfeld. Such a demonstration would be important in extending the range of validity of the principle of specific neuron encoding.

This paper offers a potentially important idea for a neurophysiological experiment which I am in no position to perform, so I am passing it along to anyone who is interested. The idea is a possibly practical way to determine the existence of *grandmother cells*—single neurons that encode object concepts (and eventually other concepts as well).

Despite all the disclaimers, what made Hubel and Wiesel's (1962) findings so exciting was the extension of Johannes Müller's (1838) doctrine of specific nerve energies to a higher level. One cannot, logically, believe that Hubel and Wiesel's papers are telling us something significant about visual coding and simultaneously assert that it is the temporal pattern of neural firing that matters or the spatial (holographic, distributed, etc.) pattern of firing frequencies over a large number of neurons in the brain. Hubel and Wiesel's research is exciting because it suggests that it is the rapid firing of a very small number of neurons that represents a line of a particular orientation at a particular location, etc.—*which* neurons are firing rapidly, not *how* they are firing or some complex firing-rate function defined over all neurons in the brain or visual cortex. Of course, there is some encoding redundancy (more than one neuron responds vigorously to any given stimulus), and each neuron has a modest generalization gradient of response to suboptimal stimuli. These are factors of some importance, but they should not obscure the basic principle of specific neuron coding of the basic line (and angle?) constituents of visual patterns. Certainly one should pay no attention to the pseudosophisticates who always darkly hint that things "can't be as simple as that" without saying why not and without offering any specific alternatives. They pretend to know something every knowledgeable person should know about why simple specific-neuron encoding cannot work. The listener often does not know, but is embarrassed to reveal his "ignorance" (often even to himself). The truth is that

specific-neuron coding will work for higher level pattern and concept representation in the nervous system; there is lots of evidence that the nervous system does work this way and no good evidence that it works any other way. There is lots more evidence to acquire and lots more thinking to do about exactly how the mind uses specific element encoding in perception, memory, cognition, response, etc. Furthermore, no one should deny the usefulness of any scientist developing precise alternative theories of coding—distributed, holographic, or whatever. But specific-neuron coding is the dominant and most plausible theory of coding in the nervous system and we should not be ashamed to admit we believe in it, if we do.

The current frontier in the long drawn-out conquest of mental phenomena by the theory of specific-neuron encoding is learned concepts. Is "grandmother" encoded in one's mind by means of a "granny" cell? Elsewhere, I have disposed of supposedly "logical" objections to the theory of learned concept neurons and specified a plausible selectional mechanism by which an unspecified (free) neuron can come to be specified (bound) to stand for a conjunction of constituent neurons (Wickelgren, 1969, Note 1). The basic principle is that each cortical neuron be weakly connected to about 10^4 other cortical neurons [approximately equal to the number of synapses/neurons in the cortex (see Pakkenberg, 1966)]. Whenever one wishes to chunk some set of attributes to specify a new chunk neuron, one inhibits the already bound neurons and primes the free neurons so that the maximally activated neuron will be one of the free neurons. The free neuron which will be most strongly activated is the one which receives the greatest input due to the convergent (direct or indirect) connections from previously bound feature neurons representing the attribute constituents of the chunk. This is the new chunk neuron representing that set of constituents. Since this neuron is the one most strongly activated following the activation of the constituent neurons, by the familiar (though unproved) neural contiguity conditioning mechanism, the synapses linking the constituent neurons to the chunk neuron are strengthened.

This work was supported by Grant BNS-03145 from the National Science Foundation.

Now we assume that strengthening one set of synaptic inputs to a neuron represses the other synapses. This "protects" the new chunk neuron from interfering input, so it cannot come to stand for any other chunk (set of constituents) in the future (at least not until and unless the previously facilitated synapses become weak through disuse or whatever).

So far we have only specified a mechanism for achieving chunking, the conjunctive aspect of concepts. There is undoubtedly also a disjunctive aspect (the sets of visual cues activating the cat concept are totally different from the sets of auditory cues, etc.) The disjunctive aspect can be handled, too, but it is beyond the scope of this brief paper.

The main point of this note is to suggest one way to answer another objection to specific-neuron encoding of concepts: that there is no (systematic) evidence for such neurons. Of course, Thompson, Mayers, Robertson, and Patterson (1970) found single neurons encoding number concepts such as two, five, six, or seven, but small numbers might be special, innately coded concepts, not representative of concepts in general. Furthermore, to my knowledge, no attempt has been made to replicate these findings. Then there is the famous "monkey hand" cell discovered by Gross, Bender, and Rocha-Miranda (1969); virtually no one expects to replicate this sort of finding because the odds that any given neuron will represent a monkey hand must be very low, if the specific-neuron encoding principle holds for concept representation. This is the essence of the problem. With so many possible learned concepts, how could a neurophysiologist have much hope of finding what concept a particular cell represents, assuming that specific-neuron encoding is true?

The answer may be to use the restricted rearing method employed so successfully in recent years to investigate the beautiful interaction of genetic and learned specification at the featural levels of the visual system. A new-born animal could be reared in an environment which was completely dark except at times when the animal was restrained, wearing a collar, and the visual field was a Ganzfeld except for a single object selected from a very small set of objects (on the order of 4 to 20). Presumably, such visual experience would be intrinsically salient enough to promote perceptual learning of visual concepts, but if not, the objects could be differentially associated with subsequent presentation (in the dark) of food, water, shock, etc.

If perceptual learning proceeds by specification of a single neuron to represent each object concept, then subsequent probing of inferotemporal cortex, association areas, etc. with microelectrodes should find cells that respond primarily to presentation of one of the objects and much less to any other familiar or unfamiliar object. Of course, since the visual experience of these animals was so restricted compared to, for example, tactual, kinesthetic, motor, and auditory experience, one expects to find fewer areas of the brain (or fewer

neurons in any area) that can be driven at all by visual stimuli. However, this should be a relatively minor problem compared to trying to figure out what combination of visual features drives a cell optimally. It is this latter problem that restricted rearing ought to make enormously easier.

Rather than using just any small set of objects, it might be best at first to use sets defined so that objects differ on two, three, or four dimensions with two, three, or four values each. Then it could be definitely demonstrated that a single cell responded in a super-additive manner to a conjunction of features compared to its response to objects possessing only a subset of the features defining the concept.

Since an object concept neuron would almost surely be activated by presentation of the appropriate object at any distance and angular orientation, there is probably no need to control this carefully, but one could if necessary. Eventually it would be interesting to determine the necessary and sufficient conditions for integrating different views (e.g., front and back) of an object into a common concept. To do this, it might be necessary to use tachistoscopic presentation of one view for a variable period of time, followed by a different view after a variable interstimulus interval.

Looking even farther down the road, one might present objects performing characteristic movements (moving up, expanding, vibrating back and forth, etc.) and see if one could find action concepts. Two objects might be presented at the same time in characteristic relations (one above the other, touching, etc.) to see if single neurons encode relational concepts. Finally, one set of objects could always be presented together and never with any of the members of another set and vice versa. Possibly there would be context neurons that responded vigorously to any object in one set, but not to any object in the other set.

Usually research on the development of some psychological competence or neural system follows research on this competence or system in adults. In the case of concept neurons, there may be good reason to reverse this order.

After writing this brief paper I learned that Michalski, Kossut, and Zernicki (1975) and Zernicki and Michalski (1974) had used visual deprivation in young kittens with selective exposure to certain objects to study the effects on units in Areas 17, 18, and 19 of visual cortex. Their experiments gave little support to the hypothesis that these visual cortical areas contain object concept neurons. However, based on receptive field studies in normal animals, there is no good reason to believe these early visual cortical areas are encoding complex, learned object concepts. Rather, these areas appear to encode the simpler, largely innately specified, line and angle constituents of object concepts. Zernicki and Michalski looked in the wrong place. Their experiment also differed in a number of other ways from that proposed here.

REFERENCE NOTE

1. Wickelgren, W. A. *Consolidation and forgetting* (Chapter 10). Unpublished chapter of book in preparation.

REFERENCES

- GROSS, C. G., BENDER, D. B., & ROCHA-MIRANDA, C. E. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*, 1969, **166**, 1303-1306.
- HUBEL, D. H., & WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 1962, **160**, 106-154.
- MICHALSKI, A., KOSSUT, M., & ZERNICKI, B. Single-unit responses to natural objects in area 19 of cats with different early visual experiences. *Acta Neurobiologiae Experimentalis* (Warsaw), 1975, **35**, 77-83.
- MÜLLER, J. *Elements of physiology* (Vol. 2). (W. Baly, trans.). London: 1842. (Originally published, 1838). Relevant section reprinted in R. J. Herrnstein & E. G. Boring (Eds.), *A source book in the history of psychology*. Cambridge, Mass: Harvard University Press, 1965. Pp. 26-33.
- PAKKENBERG, H. The number of nerve cells in the cerebral cortex of man. *Journal of Comparative Neurology*, 1966, **128**, 17-20.
- THOMPSON, R. F., MAYERS, K. S., ROBERTSON, R. T., & PATTERSON, C. J. Number coding in association cortex of the cat. *Science*, 1970, **168**, 271-273.
- WICKELGREN, W. A. Learned specification of concept neurons. *Bulletin of Mathematical Biophysics*, 1969, **31**, 123-142.
- ZERNICKI, B., & MICHALSKI, A. Single-unit responses to natural objects in visual areas 17 and 18 of cats reared under different visual experiences. *Acta Neurobiologiae Experimentalis* (Warsaw), 1974, **34**, 697-712.

(Received for publication May 31, 1977.)

ERRATUM

Newman, S. E., and Frith, U. Encoding specificity vs. associative continuity. *Bulletin of the Psychonomic Society*, 1977, **10**, 73-75. Page 75, column 1, line 26 should read: "This may have occurred since (1) in the Thomson and Tulving experiment the instructions prior to the recall test were shorter for the no-cue than for the strong-cue group and (2) strong-cue subjects, in both their experiment and ours, faced with a list of words that had not previously occurred in the experiment, . . ."