What We Have and Haven’t Learned

April 29, 2016

 In psychological science, as in other sciences, it is difficult to establish broadly consequential conclusions beyond reasonable dispute. It therefore sometimes seems that we make no progress at all; however, when I look back on what we have accomplished over the last half-century in the fields in which I have been active, I see the development of broad consensuses on several highly consequential conclusions. In this, my final column, I report what I think some of those conclusions are, and I discuss why they are broadly consequential.

First, there has been broad acceptance of the computational theory of mind. How the brain computes is fiercely disputed, but the idea that it does in some meaningful sense compute is widely accepted. Moreover, David Marr’s thesis that a computational analysis is a necessary intermediary in linking behavior with neurobiology has become widely accepted. According to Marr’s thesis, one must specify the computation that underlies a behavioral or cognitive phenomenon (e.g., the perception of surface color), the algorithm that implements it, and, ultimately, the physical implementation of that algorithm in the brain. This practice is not how behavioral neuroscience was conceived when I was in graduate school. Back then, behavioral neuroscience was called physiological psychology. There was no cognitive neuroscience, because the field of cognitive psychology, with its commitment to the computational theory of mind, was only then aborning.

Second, we have learned from behavioral experiments that foundational abstractions such as space, time, number, and probability play fundamental roles not only in our own mentation but also in the cognition and behavior of animals that we thought had no minds at all — rodents and insects, for example. We have learned from neuroscience experiments that signals based on these abstractions — spatial- and temporal-location signals, for example — are seen in individual neurons in very small brains. The number of neurons in the brain of a typical insect is about the same as the number in one voxel of a human functional magnetic resonance image (fMRI). Thus, both behavioral data and neurobiological data have taught us that it does not take a human-size brain to compute locations in space and time, to count, or to estimate uncertainty. Nor does it take extensive experience; many insects live only a few days to a few

weeks, and rodents already display behavior based on these abstractions when they are at most a few months old.

These conclusions put us at a large intellectual remove from the days when we attempted to explain even human behavior based on observed — or presumed — stimuli acting on specified receptors by means of traceable pathways between the affected receptors and the respondent muscles. In his justly famous work *The Integrative Action of the Nervous System*, Charles Scott Sherrington wrote, "From the point of view of its office as the integrator of the animal mechanism, the whole function of the nervous system can be summed up in one word, *conduction*" (1947, p. 9). This remained the prevailing view when I was a graduate student, at least in physiological psychology, although it was beginning to be challenged by the newborn field of cognitive psychology, influenced as that field was by the newly emerging field of computer science.

In those days, at least at Yale, we knew nothing about the role of dead reckoning in animal navigation (computing your present location by summing over every movement you have made since starting your trip). We also did not know that it depends on the animal's having learned the local solar ephemeris, the change in the compass direction of the sun as the day progresses. Dead reckoning reflects the mathematical fact that position is the integral of velocity with respect to time. Dead reckoning would seem to require that velocity — itself a daunting abstraction from simple sensations — be explicitly represented. That computational consideration raises the algorithmic question of whether position is represented in Cartesian coordinates (x and y) or polar coordinates (direction and distance) — or even perhaps in some more exotic form. And the integration part of the computation raises the implementational question of how nonleaky integration can be physically implemented in nervous tissue. The integration must not be leaky, because the net displacement specified by the integrator must be well and truly the arithmetic sum of all the minidisplacements over the course of a journey that may last half an hour or more and follow a tortuous course. The local solar ephemeris is the compass direction of the sun's azimuth as a function of the time of day at that latitude and that season. These are the kinds of computational, algorithmic, and implementational considerations in Marr's take on the computational theory of mind and its role in behavioral and cognitive neuroscience. The notion of conduction is not adequate even to pose these questions, let alone to offer answers to them. We have learned that we cannot understand the causation of behavior, nor how the brain works, until we know the answers to these sorts of questions.

Behavioral neuroscientists have underplayed the astonishing level of abstraction that we see in the signals from individual neurons. Take for example, the head-direction cells; they are tuned to the compass direction in which the head is oriented. That is, these individual neurons fire selectively when the head points north, or northeast, or southwest, depending on which neuron one records from. Mathematically, compass direction is a point at infinity. It picks out an infinite family of polarized parallel lines — all those polarized parallel lines that, when followed to infinity in their positive directions, intersect at that one point. A head-direction cell tuned to a compass direction fires maximally whenever the midline of the head forms a cell-specific angle with a line in that family. In the general case, there are no compass-direction stimuli in the experienced environment, no sensible parallel rays emanating from fixed points on the compass. Estimating compass direction requires environmental input, to be sure, but its dependence on sensory input can be understood only in computational terms. These terms often challenge our patience for mathematical exposition — an exposition that would, in any event, take much more space than I have here.

What we haven’t yet learned are the answers to the computational questions that we have learned to ask. We do not yet know how the brain implements the basic elements of computation (the basic operations of arithmetic and logic). We do not yet know the mind’s computational primitives. We do not yet know in what abstract form (e.g., analog or digital) the mind stores the basic numerical quantities that give substance to the foundational abstractions, the information acquired from experience that specifies learned distances, directions, circadian phases, durations, and probabilities. Much less do we know the physical medium in nervous tissue that is modified in order to preserve these empirical quantities for use in later computations. Already as an undergraduate, I wanted to know the physical basis of memory in the brain. I begin to think that we are not to know this in my lifetime, but science often progresses in sudden and unexpected spurts, so I still hope to know it. œ