projectivism by suggesting that values have the same real, i.e., nonsubjective, status that other secondary qualities have, viz., a power to produce certain effects in us. The same maneuver might be extended to the unity problem. Projectivism seems to be a kind of eliminativism. But a secondary–quality view, as McDowell points out, preserves a legitimate ontological niche for the entities in question, indeed, a *not–purely–subjective* ontological niche. For our topic, it would say that the property of entitivity is the ("objective") property of having the power to produce certain effects in human beings. As such, entitivity is secure, even in a scientific image of the world.

Many questions might be raised for the secondary–quality gambit. But suppose for the sake of argument that it succeeds. This might defeat projectivism, but it does not defeat the suggestion that cognitive science can have ramifications for metaphysics, even revisionary metaphysics. This is because the conclusion that entitivity is a secondary quality is already a metaphysical conclusion, plausibly a revisionary metaphysical conclusion. So the kind of connection I have adumbrated between cognitive science and metaphysics would still be exemplified.

ALVIN I. GOLDMAN

University of Arizona

EPISTEMOLOGY IN THE AGE OF NEUROSCIENCE*

IN a recent archaeological dig through several decades of my papers, I unearthed the graduate student stratum. At that level, I discovered a dusty bundle of final examination papers from Oxford, and, in pondering the test questions, I found myself viewing most of them as old curiosities. Among such curiosities for me are the Gettier problem, the nature of sense data, the nature of incorrigible foundations of knowledge, and the constituents of the corpus of a priori knowledge.

Though the examination papers are in fact less than twenty years old, so much has changed, and changed profoundly, in the field, that it may not be an exaggeration to say that we are in the midst of a

* To be presented in an APA symposium on Epistemology and the Philosophy of Mind, December 28, 1987. Alvin I. Goldman will be co–symposiast, and George Bealer will comment; see this JOURNAL, this issue, 537–544 and 553–555, respectively, for their contributions.

Grateful thanks are owed to Paul Churchland, Patricia Kitcher, Stephen P. Stich, Terrence J. Sejnowski, and Warren Dow for discussion and advice. The three figures below are adapted from Paul Churchland's *Matter and Consciousness*, 2nd ed. (Cambridge, Mass.: MIT, forthcoming).

0022-362X/87/8410/0544$01.00

paradigm shift. Since one cannot accurately speak for the whole field, I should properly speak only for myself (and Paul Churchland). With that proviso, it seems to me that the general frame of reference within which we might hope to discover how humans learn, understand, and perceive is undergoing a major reconfiguration. Most of the questions which used to preoccupy us as graduate students and whose answers seemed necessary to advancing the general program in epistemology, now look either peripheral or misguided, and the general program itself looks troubled.

I am no longer preoccupied with the nature of absolute foundations, because it does not look as if there are any; or with a priori knowledge, because there probably is not any, or with sense data, because that is a mixed–up way of thinking about sensory processing. It is doubtful that knowledge in general is sentential; rather, representations are typically structures of a quite different sort. Whatever reasoning and information processing turn out to be, formal logic is probably not the model, save perhaps for a small part. Decision theory, confirmation theory, the predicate calculus, etc., beautiful and magnificently clever though they are, do not appear destined to play a central part in the theory of how, in fact, human and other nervous systems solve problems and figure things out. Inductive logic does not exist, and does not show any positive signs in that direction; 'inference–to–the–best–explanation' is a name for a problem, not a theory of how humans accomplish some task.

Formal semantics now looks like a thoroughly misbegotten project which cannot even begin to explain how human language is meaningful. Hilary Putnam[1] has provided grounds for this view, and empirical data from linguistics also seriously undermines the general idea.[2] Formal semantics may be invaluable for certain purposes, but, for generating a theory of meaningfulness in human language, it looks like a dismal failure. Surely there is something bizarre about the idea that a theory of meaning that has nothing whatever to do with human psychology or neurophysiology can explain the meaningfulness of language and how representational structures relate to the world. And what of truth? If representational structures are not sentences (propositions), they are not truth-valuable; if they are to be evaluated, it must be in some other way.[3] Consequently, the very concept of truth appears to be in for major reconsideration.

[1] *Reason, Truth, and History* (New York: Cambridge, 1981).

[2] George Lakoff, *Women, Fire, and Dangerous Things* (Chicago: University Press, 1987).

[3] For a further discussion of the deep problems with truth and rationality, see Stephen P. Stich, *The Fragmentation of Reason* (Cambridge, Mass.: MIT Press, forthcoming).

Thus, knowledge and belief, reference, meaning, and truth, and reasoning, explaining, and learning, are each the focus of eroded confidence in "the grand old paradigm," a framework derived mainly from Logical Empiricism, whose roots, in turn, reach back to Hume, Locke, and Descartes. This is certainly not to say that nothing has been achieved, or that everything worked out from within the confines of the old assumptions is bunk. On the contrary, there are surely many enduring results, although at this stage I find it difficult to know how to tell the enduring from the ephemeral.

Moreover, it is not that there has been a decisive refutation of "the grand old paradigm." Paradigms rarely fall with decisive refutation; rather, they become enfeebled and slowly lose adherents. Confirmed practitioners can always continue, secure in the faith that a new wrinkle may yet satisfy the critics. But many of us sense that working within "the grand old paradigm" is not very rewarding. By contrast, there is considerable promise in a naturalistic approach, which draws upon what is available in psychology and neuroscience to inform our research. There are remarkable new developments in cognitive neurobiology which encourage us to think that a new and encompassing paradigm is emerging. Epistemology conceived in this spirit is what W. V. Quine[4] has called *naturalized* epistemology.

THE BIOLOGICAL PERSPECTIVE

The fundamental epistemological question from Plato onward is this: How is it possible for us to represent reality? How is it that we can represent the external world of objects, of space and time, of motion and color? How do we represent our inner world of thoughts and desires, images and ideas, self and consciousness? Since it is, after all, the nervous system that achieves these things, the fundamental epistemological question can be reformulated thus: *How does the brain work?* Less cryptically and more accurately, the question is: *How, situated in its bodily configuration, within its surrounding physical environment, and within the social context it finds itself, does the brain work?* Answers here will be descriptive first and foremost, but the normative dimension of epistemology enters when we can draw on the descriptive basis to compare and evaluate styles of computation and representation, and determine how to improve upon particular computational and representational strategies. Once we understand what reasoning is, we can begin to figure out what reasoning *well* is.

This is a good time to be naturalizing epistemology, and three factors in particular indicate this. First, technological developments

[4] "Epistemology Naturalized," *Ontological Relativity and Other Essays* (New York: Columbia, 1969), pp. 69–90.

in research in the neurosciences during the past twenty years have been spectacular, and a truly impressive amount is now known about the microstructure and micro-organization of nervous systems. For example, we are learning in detail about the pathways for particular neuron types in the visual system, about the physiological properties of different types of neurons, and about the different tasks performed by distinct neural populations. Techniques for addressing nervous systems at a variety of levels of organization have revealed anatomical and physiological data that suggest that the time is ripe for genuine theorizing about how macro effects are the outcome of neuronal properties.[5]

Second, the advent of cheap computing makes it possible to simulate neural networks and, hence, to investigate computational properties at the *circuit* level. Since we do not have physiological techniques to address this level in actual nervous systems, the computer simulation of biological circuits is a crucial adjunct to available neuroscientific techniques. Cheap, fast computing is essential for simulation, because nervous systems have a parallel architecture, and huge numbers of computational events are going on simultaneously in the network. Nevertheless, even new computing technology such as the current-generation connection machine[6] is still far inferior to the human brain in computing capacity. The communications bandwidth of the connection machine is about 10^{10} bits per second, which is extraordinary. But the *average* communications bandwidth used by the human brain is about (10^{11} neurons)(5×10^3 connections/neurons)(2 bits/connection/sec), which means that the brain processes on average 10^{15} bits/second.[7]

Third, studies of animal behavior in ethology, psychology, linguistics, and clinical neurology have become increasingly sophisticated and have yielded crucial and often surprising data concerning the capacities of nervous systems. Within clinical neurology, one of the most important discoveries concerns the multiplicity of memory systems in humans and in monkeys.[8] Within psychology, the discoveries that virtually all categories show prototype effects[9] and that images

[5] See my *Neurophilosophy* (Cambridge, Mass.: MIT Press, 1986).

[6] W. Daniel Hillis, *The Connection Machine* (Cambridge, Mass.: MIT Press, 1985).

[7] Terrence J. Sejnowski, review of *The Connection Machine, Journal of Mathematical Psychology,* XXXI (1987): in press.

[8] This is thoroughly and clearly discussed by Larry Squire in *Memory and Brain* (New York: Oxford, 1987).

[9] Eleanor Rosch, "Human Categorization," in N. Warren, ed., *Studies in Cross-cultural Psychology* (London: Academic Press, 1977).

are prevalent in information processing[10] stand out as particularly important.

Adopting the biological perspective, we find that a number of points have special significance. Not that these points are generally unknown, but from a biological perspective, they acquire distinctive salience, for they shape the way we think about the problems of vision, learning, spatial representation, sensorimotor control, and so forth.

The most fundamental point is that the human brain is a product of evolution. This is worth reflecting on for three reasons: (1) In many important respects, the human brain is very similar in structural components and organization to other primate brains, and not very different from other mammalian brains. Cortical organization, for example, seems generally shared among mammals. Our brains share some basic properties with even the simplest nervous system. Like ours, mammalian nervous systems are networks of connected units; these units, neurons, function basically in the same way in all nervous systems; we share the same neurochemicals. There are bound to be some differences, at least because the human communication capacity appears to be distinctive, but this is probably owed not to a special "wonder lobe" but to subtle, fine–grained wiring differences. The anatomical and physiological similarities are important, because they mean that we should expect that how *we* learn, remember, see, hear, and respect, is not fundamentally different from how other organisms do those things.

(2) Cognition is not neatly detachable from the organism's ecological niche, way of life, and bodily structure. Nervous systems are not general–purpose computers. They have evolved to accomplish a certain range of tasks, and the architecture supports those tasks. There is a fatal tendency to think of the brain as essentially in the fact–finding business—as a device whose primary function is to acquire propositional knowledge. At its best, supposedly, it discovers truth–for–its–own–sake. From a biological perspective, however, this does not make much sense.

Looked at from an evolutionary point of view, the principal function of nervous systems is to enable the organism to move appropriately.[11] Boiled down to essentials, a nervous system enables the organism to succeed in the four F's: feeding, fleeing, fighting, and reproducing. The principal chore of nervous systems is to get the body

[10] Stephen Kosslyn, *Image and Mind* (Cambridge, Mass.: Harvard, 1980).

[11] J. L. McClelland and D. E. Rumelhart, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Cambridge, Mass.: MIT Press, 1986).

parts where they should be in order that the organism may survive. Insofar as representations serve that function, representations are a good thing. Getting things right in space and time, therefore, is a crucially important factor for nervous systems, and there is often considerable evolutionary pressure deriving from considerations of speed. Improvements in sensorimotor control confer an evolutionary advantage: a fancier style of representing is advantageous *so long as it is geared to the organism's way of life and enhances the organism's chances of survival.* Truth, whatever that is, definitely takes the hindmost.

(3) We cannot expect engineering perfection in the products of evolution. Improvements in nervous systems are not built by starting from scratch, even though that might yield the best design. They are modifications of structures and patterns that already exist. Evolution has to achieve improved organization by managing with what is there, not by going back to redesign the basics. As we consider research strategy, this is an important consideration. It means that, if we approach the problems of nervous–system function as strictly engineering problems, setting our goals in terms of how a task (for example, stereopsis) could *in principle* be done, we may find a cunning engineering solution which is nothing like the solution the brain has found. In framing hypotheses concerning brain function, it will be essential to consider constraints not only at the behavioral level, but also at the neurobiological level. Unless we go *into* the black box, we run the considerable risk of wasting our time exploring remote, if temporarily fashionable, areas of computational space.

WHY CONNECTIONISM IS IMPORTANT

If representational structures are not sentence–like, what are they? If computation is not logic–like, what is it like? And how can we find out? If formal semantics is the wrong approach to explaining meaningfulness, what would work? These seem to me to be the central questions, or at least some of the questions relevant to epistemology. Connectionism (also known as Parallel Distributed Processing, or PDP) is important, because it constitutes the beginnings of a genuine, systematic alternative to the "grand old paradigm." It appears to have the resources to provide neurobiologically plausible answers to these central questions (see McClelland and Rumelhart, *op. cit.*).

Connectionist models illustrate what representations might really be, if not sentence–like, and what neurobiological and psychological computation might really be, if not logic–like. They free us from the conviction that the sentence/logic model is inevitable. Moreover, the design of the models is inspired and informed by neurobiology; so they are more biologically realistic than sentence/logic models.

A connectionist model is characterized by three architectural elements: (1) processing units, (2) connections between processing units, and (3) weights, which are differential strengths of connection between processing units. The processing units, like neurons, communicate with each other by signals (such as firing rate) which are numerical rather than symbolic. In typical models, the units are arranged in layers: an input layer, an output layer, and, intervening between, the layer of so–called "hidden units" which are fully connected to the input and output layers (see Figure 1). The processing units sum the inputs from the connections with other processing units, each input weighted by the strength of connection. The output of each processing unit is a real number that is a nonlinear function of the linearly summed inputs (see Figure 2). The output is small when the inputs are below threshold, and increase rapidly as the total input becomes more positive. To a first approximation, the activity level can be considered the summed postsynaptic potentials in a neuron, and the output can be considered its firing rate.

What is remarkable about such systems is that they can be trained to learn highly complex input–output functions. This learning is not merely a look–up table achievement, because the small number of hidden units makes that impossible. Moreover, the system can gener-

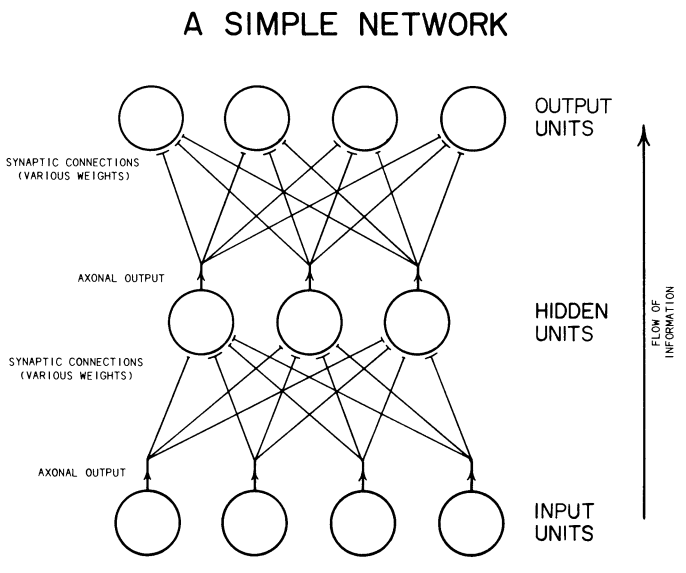

Fig. 1: Schematic model of a three-layered network. Each layer is fully connected to the layer adjacent to it.

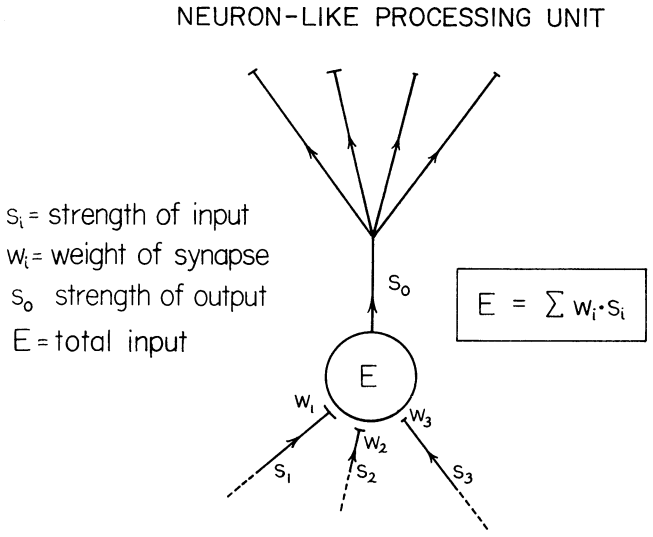

Fig. 2. Schematic model of a processing unit receiving inputs from other processing units.

alize to new cases, and it can make sensible approximations in early training or when there is no correct answer.

NETtalk[12] is perhaps the most spectacular example of a connectionist network. It learns to convert written text to speech. The network does not have any initial or built–in organization for processing the input, or, more exactly, for mapping letters onto sounds. All such organization emerges during the training period. The values of the weights are determined by using the learning algorithm called "back–propagation of error."[13] The strategy exploits the calculated error between the *actual* values of the processing units in the output layer and the *desired* values, which are provided by a training signal. The error signal is propagated from the output layer backward to the input layer and is used to adjust each weight in the network. The network learns as the weights are changed to minimize the mean squared error over the training set of words. Thus, the system can be characterized as following a path in weight space until it finds an error minimum (see Figure 3). The important point to be illustrated,

[12] Terrence J. Sejnowski and Charles R. Rosenberg, "Parallel Networks that Learn to Pronounce English Text," *Complex Systems,* I (1987): 145–168.

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," in McClelland and Rumelhart, eds., *op. cit.*

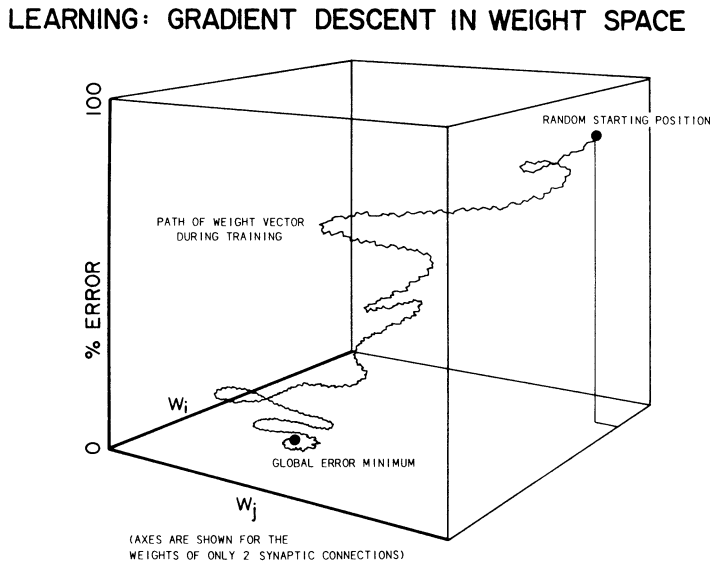

Fig. 3. Schematic drawing of a path followed in weight space as the network finds a global minimum.

therefore, is this: the network processes information by nonlinear dynamics, not by symbol manipulation and rule following. It learns by gradient descent in a complex interactive system, not by generating new rules. Representations in the hidden units turn out to be not symbols, but patterns of activation (see Sejnowski and Rosenberg, *op. cit.*).

The general strategy in connectionism is to model information processing in terms of the trajectory of a complex, nonlinear dynamical system in a very high dimensional space. This structure does not resemble sentences arrayed in logical sequences, but it is potentially rich enough and complex enough to yield behavior that can support semantic and logical relationships.

The connectionist approach is still very new, and many questions remain unanswered, but progress so far is impressive. Some models address specific neurobiological problems and are highly constrained by neurobiological data. Others address a higher level of organization. At this stage, the important thing is that a model suggest experiments, preferably at the neurobiological level, where the results will suggest further modifications in the model. Additionally, it is now imperative that additional information-processing algo-

rithms be devised. Although it may be too soon to say whether any existing models successfully capture how information processing is actually accomplished in nervous systems, the general approach has the right character in so many dimensions that it has to be taken seriously.

The breakthroughs in network modeling, together with new discoveries in neuroscience and psychology, suggest that it really is possible to understand the fundamental principles governing brain function and, thus, to understand the nature of representing and reasoning. It is also clear that finding solutions to these problems is an inescapably interdisciplinary task, requiring networks of researchers: neuroscientists, modelers, ethologists, psychologists, linguists, and philosophers. My hunch is that epistemology will never look the same.

PATRICIA SMITH CHURCHLAND
University of California/San Diego

THE BOUNDARY BETWEEN PHILOSOPHY AND COGNITIVE SCIENCE*

The main issue before us concerns the boundary between philosophy and empirical science. Patricia Smith Churchland and Alvin I. Goldman advocate an aggressive empiricist philosophy according to which many fundamental philosophical questions lie within the province of empirical science—specifically, cognitive psychology, physiology, and perhaps experimentally based computer science. The primary problem (not to say the only problem) with this aggressive empiricism concerns modality. Like the answers to questions in pure mathematics, the answers to basic philosophical questions are *necessary if true*. For example, if justified true belief is not knowledge, then necessarily justified true belief is not knowledge. Even if all and only cases of justified true belief were, in fact, cases of knowledge, that would not show that justified true belief is knowledge; for the mere possibility of a case of justified true belief that is not knowledge suffices to show that justified true belief is not knowledge.

* Abstract of a paper to be presented in an APA symposium on Epistemology and Philosophy of Mind, December 28, 1987, commenting on papers by Alvin I. Goldman and Patricia Smith Churchland, this JOURNAL, this issue, 537–544 and 544–553, respectively.

0022-362X/87/8410/0553$00.50