

Patrick Suppes

RESPONSE TO CLAUDIA ARRIGHI

Arrighi has, in her variety of remarks and analyses of my work, caught very well the two most salient features, namely, on the one hand, my interest of long standing in formal axiomatic problems concerning scientific theories, and on the other hand, in sharp contrast my interest in experiments and the detailed analysis of experimental data. Among philosophers I am probably better known for my formal interests than my experimental ones. So in this response I will have a good deal to say about the experimental side of my own thinking and what I believe to be the proper role of the detailed consideration of experiments in the philosophy of science.

Arrighi has a number of different and interesting things to say about behaviorism, some general, some specific to my own views. I will begin with a number of comments about behaviorism. They will very much reflect the changes in my own views as I began to work on the brain in 1996. The first thing to note is how unsatisfactory from a formal standpoint the definitions of behaviorism are – this includes my own, of course. I don't mean to suggest by this that we should not discuss what we think behaviorism is but rather, that we recognize from the beginning that we will not end up with some satisfactory detailed systematic idea. In this connection the conceptual importance of whether or not to include the brain in discussions of behaviorism is paramount. I am reminded of the contrast

between theories of matter before and after atomism was finally accepted as the correct theory of matter at the beginning of the twentieth century. When I say “correct theory of matter” I mean the correct theory in terms of substantial data available at that time. Moreover for the experiments supporting the existence of atoms, the correctness of the periodic table for the elements, etc. represent experiments that are not false. They stand for all time as a remarkable achievement in the same way as Ptolemy’s astronomy did for 1500 years. They are approximately, correct as much of Ptolemy’s astronomy still is. This doesn’t mean further improvements don’t take place. It’s important, from my standpoint, to have a conception of science, including behavioristic varieties, that have a place for correctness at a certain level of approximation or coarseness, as well as obvious and continued improvements in what had gone before. This seems to be even more true, looking forward, of what we should expect in psychology than it is in physics, where so much has already been accomplished, even though the horizons of physics now seem unbounded in terms of what we can hope to learn over the next several hundred years. Anyway, the introduction in the twentieth century of substantial concentration on the brain has changed psychology for ever. Even though that change has been taking place, – to be conservative about the beginning –, since the excellent discussion of what was known at the time about the brain in the second chapter of William James’ *Principles of Psychology* (1890) to the modern focus on brain imaging. This comes in four important varieties: electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), and functional nuclear magnetic resonance imagery (fMRI) – in the usual biological and medical literature the word nucleus, “nuclear” or an abbreviation for it are omitted because of the fear that medical patients may be disturbed by a form of imaging that refers to nuclear activity. That these methods all represent triumphs of the twentieth century are well known. The lateness of nuclear magnetic resonance is reflected in the fact that the fundamental physics was only fully worked out in the 1940s, for which Bloch and Purcell later received a Nobel Prize. Even MEG only came on the scene in the 1970s be-

cause of the availability of superconducting quantum interference devices to receive the extremely weak magnetic signals from the brain. So brain imaging is something new, and the current torrent of activity is something really new. This doesn't mean that there weren't excellent experiments on EEG at a much earlier time, beginning indeed with the early work of Hans Berger starting in 1910.

Well after this review of the growth of brain imaging, the natural question is this. Does the behavior of our brains count as a part of behaviorism? It is certainly true that the classical definitions do not include behavior of our brains, but this is only because at that time so little work had been done, even though some of it in the 1940s and the 1950s was superb. Moreover, various kinds of behavioristic experiments not concerned with the brain were focused on measuring nonobvious responses. Some examples are: Galvanic skin response, heart beat, and careful measurements of the latency of responses, that is, speed of responses, in given situations. So it is natural, it seems to me, now as we get to work in the twenty-first century, to extend behaviorism to include behavior of the brain. The real battle among philosophers will therefore be between those who think that mental activity is just a form of activity of our brains, a naturalistic thesis about the physical nature of the mind as opposed to those who find in the concept of mind something that goes beyond the possibility of a physical account. Obviously I squarely belong to the former, and not the latter, school of thought. Keeping these remarks about the brain in mind, it is clear that many of my own earlier writings about behaviorism need to be modified.

So let me turn to the contrast in the 1969 article on the vocabulary of behaviorism versus that of non-behaviorism. I put in this latter category, as Arrighi quotes, such words as "intention, belief, purpose of behavior, rule-following behavior". Much more than in 1969, in fact I would say very much more so, I now want to include intention as a concept on the behavioral side. Moreover, I want to start by characterizing, in a way close to Aristotle in the *De Anima* but with a slightly different flavor, life as animate matter. It is then a feature of animals among living creatures, and therefore a feature of much animate matter, to have intentions. This is not the place to

go into a long argument about why the usual concerns about determinism are misplaced. I have decided views on this and have published them elsewhere (Suppes [1988], [1991], [1993]). Here I'll just take it as a fact that I have an inclusive concept of behaviorism that's happy with intentions and purposes, which, as Aristotle certainly held are just as much a part of the natural world as familiar properties of physical objects. A response to this might be said, well, such a catholic and all-inclusive behaviorism can hardly be rejected by those who believe in mental activity and mental concepts. But my answer is that I am excluding a variety of philosophical thought, namely, all that which is concerned to separate the mental from the physical. A good example would be mental representations. I don't really know what a mental representation is unless what is meant by it is a brain representation, for example, what is a mental representation of the word "isomorphism"? (For more details on my conception of mental representation, see the last section of Chapter 3 of Suppes [2002]).

After that long remark, let me make a minor one about the nature/nurture controversy. It is inevitable, I suppose, that behaviorists are thought to always favor the nurture side of the debate. This is certainly wrong about me and even more so now that I have become much more absorbed in the behavior of the brain and in the transformations of perceptions as they move from the peripheral nervous system to the cortex. Anyone who thought wholly in terms of nurture is surely ignorant about the basic physics of what is going on. Perhaps the thing that is most important, and at the same time most neglected in the discussions of this controversy, is the fact of the enormously complicated physical transformations from every kind of perception essentially into an electrical signal going from the peripheral nervous system to the brain, especially to the cortex. To think for a moment that the intricate machinery of either the auditory or visual system, to take the two most important senses for perceiving language and much else about the world, are derived from scratch as a matter of nurture would be a piece of scientific lunacy. It took millions of years for these two systems to evolve. In fact, rightly put, hundreds of millions years. After this enormously long

period of evolution, it is scarcely surprising that we find it difficult to build robots that have a comparable sense of vision or hearing. On the other hand, one of the great marvels of evolution is how flexible this electrical and chemical system is. It makes itself available for fine tuning the job of nurture not only in the matter of language but the matter of almost all other aspects of interacting with the world. I guess I've come to see the nature/nurture controversy as tedious and uninteresting. The extreme view on either side seems hopelessly wrong.

Much of what Arrighi has to say about learning models in that section of her commentary I agree with. I just want to comment upon how restricted, on the one hand, the concepts of behaviorism are and, on the other hand, how naturally the theory extends into the estimation of parameters for unobservable processes. Let me clarify by two examples. The first is about the rhetoric aimed at stimulus-response theories. In fact, in the mathematical formulations of theories of the 1950s no claim was usually made about the observability of the stimuli taken in by an organism in the experiment. The experimental *situation* was described, not the stimuli actually perceived by the subjects of the experiment, be they gold fish or humans.

So how was the concept of stimuli dealt with? Well, there was no attempt to have even a theoretical identification of what the stimuli were in all the standard theories that were given formal expression. What was introduced as a parameter was the estimation of the number of stimuli. So the variability in behavior, for example, variation of responses under various partial schemes of reinforcement from one trial to another, would be fit better by the theory by detailed estimation of the number of stimuli.

In a similar vein it was standard to estimate a learning parameter, something that is still done, but that learning parameter was not in itself directly observable, but could only be estimated from theoretical machinery assumed in some particular version of behavioral learning theory given mathematical expression. The only two notions that were left for direct observation were that of response and the reinforcements given under a particular schedule by the experi-

menter. But in detailed theories, even this notion of reinforcement was made subjective in terms of what was perceived by the subjects, as opposed to what had been selected as reinforcement by the experimenter. So the notion of observability was never one that was given rigorous expression in this whole tradition. This is not to be negative about what was done, but just to be realistic from a broad philosophical standpoint about the intricacies of using the concept of observability.

One of the formal things that I concentrated on in learning theory was the proof that familiar linear models that did not explicitly use a concept of stimulus were the asymptotic limit of stimulus sampling models, where here “asymptotic” means asymptotic in the sense that the number of stimuli approaches infinity. The intricate and detailed proof of this result is to be found in Chapter 8 of Suppes [2002], a proof that is so tedious that my own view of the main interest is in showing how difficult it is to provide absolutely complete reductions of one theory to another even in closely related areas of science. Broader theses of reduction seem hopeless in most cases of serious realization, even though it may be important, as in my own view about the mental as a form of brain activity. This explicit focus represents a fundamental shift for many in the conception of what it means to have a mental life.

I turn now to the brain experiments which have been one of my principal interests since 1996. Let me concentrate on just a few remarks, and try to avoid in the process of doing so, giving an overweening amount of detail about the experiments. First I would like to reformulate what Arrighi has to say about the way I describe the temporal data in our EEG recordings, time-locked to the presentation of a verbal stimulus. From my standpoint the brain is computing first the identification of the transformed character of the stimulus as it moves through either the visual or auditory nervous system to reach the cortex. In this process, first it is transformed into electrical currents that in themselves generate, probably mainly in the synapses of the neurons in the cortex, the electrical field that we record. So our observable data consist of recordings of either an electrical or magnetic field reflecting the electrical activity and secondar-

ily chemical activity of the brain to the reception of such transformed stimuli (I'll speak in terms of electric, I could say electromagnetic or just magnetic but since the experiments are observing the electrical field I shall stick to electrical activity). The point to start with is to emphasize how transformed this electrical activity is in comparison with the physical nature of the stimulus that entered either the rod and cones of the eye or the sound pressure wave that entered the auditory system. Looked at simply from the outside it is utterly remarkable that so much detail in the original source of the stimulus is preserved as invariant. Of course, we are still in the process of discovering these invariants.

My second point is that I am looking then for what in this electrical field that we are recording represents, in the traditional sense of representation, the original verbal stimulus, which itself was a temporal activity. I refer in the case of reading not to the inert printed word on the page, but to the physical activity of observing this word and having electro magnetic phenomenon now at the level of light entering the rods and cones of the eyes. So we are looking for a standard representation, isomorphic in the sense of structure, but remarkably and wonderfully different in external appearance. The difference is almost as great as that between empirical procedures of measurement and isomorphic representations of those procedures to abstract numerical operations on numbers, to give us standard measurements of physical properties or processes.

This return to the image of the representation of concrete experimental procedures of measurement by abstract numerical operations suggests that we could, in principle, aim for the same thing in the case of the brain. So, though it is of course desirable to have a structural isomorphism between spoken speech, for example, and the brain representations of that speech when heard, and also for the even more complicated case when that speech is produced and is initially formulated in the cortex as something to be spoken, so we could go in a different direction and try to make a very detailed science out of the middle, so to speak. Namely, we would produce a detailed abstract structure for what we think of as the mental activities of ourselves as humans, or in simpler cases, of other animals. We

would then want to establish an isomorphism between the brain representation of a spoken sentence and the mental representation of that spoken sentence where the terminology of the brain representation would be physical. In particular it would be electromagnetic in character at the bottom level or at least at the level of EEG recordings. In contrast we would have an abstract formulation in terms of what we would like to think of as the purely mental. Such a program does not seem impossible to work on, but it is quite clear that thinking in philosophy of mind or in cognitive science has in no sense moved very far toward creating such a systematic theory of mental representations. Having such a separate theory of mental representations would in no sense imply we should move away from a completely naturalistic and physicalistic attitude toward the mental, namely, that animate matter can have mental properties, just as it can have electromagnetic properties, mechanical properties, and chemical properties. The viewpoint I am expressing is most certainly not a new one. It is very close to the attitude, expressed in different terminology by Aristotle in the *De Anima* and by Aquinas in his extensive commentary.

Arrighi mentions in several different places and from different angles my insistence on the continuing importance of Hume's central mechanism of the mind, association. I want to say something more here about what appears to be the universal role of association in the brain. Already in the eighteenth century Hume gives a number of good examples of association's role in learning about the world, as well as in the development of the passions, or what we would call now the emotions. Perhaps the cleverest and deepest example of association in the treatise is the analysis at the beginning of Book II of the passion of pride. In the heyday of cognitive science between 1965 and 1980 it was customary to shrug off the mechanism of association as old-fashioned and inadequate to handle the sophisticated concepts being developed by either cognitive psychologists or analytic philosophers. We have by now returned to a sounder view of these matters. The fundamental importance of association in the neurosciences is thoroughly appreciated, and the study of learning processes in terms of association is focused on in both

neuroscience and computer science. A common and mistaken complaint, during those heady days just mentioned, was the claim that, of course, association could not give an account of such complex concepts as those involving rule learning and following. But this is simply a mistake. We now know, as one of the great foundational clarifications of the twentieth century. Rules, or in more technical terminology, any computable function, can be constructed from very simple ingredients, be they a simple Universal Turing Machine's small number of states and unlimited tape, or an associative network active roles and links. Clear and definite mathematical proofs of these matters are widely available in the literature, so I will not say more. Some clear examples of association in artificial intelligence and machine learning are to be found in papers of my own with others on machine learning of robotic language (Suppes *et al.* [1995a]; Suppes *et al.* [1995b]; Suppes and Liang [1996]; Suppes *et al.* [1966]; and Suppes and Böttner [1998]). The scheme of learning in these machine learning papers is technical and in its own way rather complicated, but is, of course, simplicity itself compared to the much more complicated constructions taking place in our brains. On the other hand, there is much consensus among those working on these matters from a mathematical and a formal standpoint in neuroscience that the two main concepts that need to be understood in terms of how they can be realized in large assemblies of neurons are the concepts of association and of memory storage and retrieval. For a quite recent view of what a realistic neural model looks like to realize these two concepts, see Valiant [2005].

There is one point about Hume connected to a quotation from Bauer and Anderson given by Arrighi that I want to mention. Hume did emphasize that complex ideas were composed out of simple ones. I look upon this as a mistake just as William James did; in fact, it was James' main dissatisfaction with Hume's theory of association (James [1890], pp. 594-604). Fortunately, Hume did not give a lot of examples of the sort that would hang him out to dry for several centuries following. He just emphasized simple ideas too much. We could, of course, take another course of simplicity in defense of Hume, but not one that he had in mind, namely, the

development of conditioning in very simple animals, such as *Aplysia*, is inevitably much simpler than the associations that are the basis of learning in humans or other higher animals. But this is not a defense of Hume, just a way of indicating how there is a natural concept of greater and less simplicity, or, if you prefer the opposite, complexity in the learning of organisms.

My final comment on Arrighi concerns her closing remarks on the debate about connectionism and cognitive architecture. Going back to the well known article of Fodor and Pylyshyn [1988], she states the alternatives perhaps too simply, but I think she catches the main point right. My answer is clear. There is extraordinarily weak positive evidence and much negative evidence that the mental processing involved in perception or in cognition do not even begin to approximate the formal systems of inference available in “classical symbolic systems.” It’s nice to dream that organisms including humans, are so constructed, with symbols at the ready. There is really almost no evidence that it is correct or could possibly be correct. Now, this is not the place to begin the argument all over again. Let me just remind the reader, however, that the most universal classical symbolic system is that of classical logic, and it is well known how little of ordinary reasoning as expressed in ordinary language can in any direct way be reduced to this formal system. I do not think anyone who has studied the problem with any carefulness really believes that, beneath the enormous problems of expressing even the measured sentences of good lawyers, some simple logical system close to the classical one is doing the cognitive work. I cannot say positively exactly how that work is done. Trying to understand in serious detail the physical procedures of natural computation used by the brain is a worthy endeavor for the future.

REFERENCES

- Aristotle [1975], *De Anima (On the soul)*, Harvard U.P., Cambridge (MA), 4th edn. English translation by W.S. Hett, First published in 1936.
- Aquinas T. [ca. 1270/1999], *A Commentary on Aristotle’s De Anima*, Translation by Robert Pasnau, Yale U.P., New Haven.

- Fodor J. and Z. Pylyshyn [1988], "Connectionism and Cognitive Architectures: A Critical Analysis", *Cognition* 28: 3-71.
- James W. [1890/1950], *Principles of Psychology*, Dover Publishing, New York, First published in 1890.
- Suppes P. [1988], "Comment: Causality, Complexity and Determinism", *Statistical Science* 3: 398-400.
- Suppes P. [1991], "Indeterminism or Instability, does it Matter?", in G.G. Brittan, Jr. (ed.), *Causality, Method, and Modality*, Kluwer Academic Publishers, Dordrecht, pp. 5-22.
- Suppes P. [1993], "The Transcendental Character of Determinism", in P.A. French, T.E. Uehling, and H.K. Wettstein (eds.), *Midwest Studies in Philosophy*, vol. 18, University of Notre Dame Press, Notre Dame, pp. 242-257.
- Suppes P. [2002], *Representation and Invariance of Scientific Structures*, CSLI Publications, Stanford (CA).
- Suppes P., M. Böttner, L. Liang, and R. Ravaglia [1995a], "Machine Learning of Natural Language: Problems and Prospects", in M. de Glas and Z. Pawlak (eds.), *Proceedings of the Second Conference on the Fundamentals of Artificial Intelligence*, Angkor, Paris (France), pp. 511-525.
- Suppes P., M. Böttner, and L. Liang [1995b], "Comprehension Grammars Generated from Machine Learning of Natural Language", *Machine Learning* 19: 133-152.
- Suppes P. and L. Liang. [1996], "Probabilistic Association and Denotation in Machine Learning of Natural Language", in A. Gammerman (ed.), *Computational Learning and Probabilistic Reasoning*, John Wiley & Sons, Ltd., Sussex (England).
- Suppes P., M. Böttner, and L. Liang [1996], "Machine Learning Comprehension Grammars for Ten Languages. *Computational Linguistics* 22: 329-350.
- Suppes P. and M. Böttner [1998], "Robotic Machine Learning of Anaphora. *Robotica* 16: 425-431.
- Valiant L G. [2005], "Memorization and Association on a Realistic Neural Model", *Neuro Computation* 17(3): 527-556.