

FUNCTIONALISM AT FORTY :

A CRITICAL RETROSPECTIVE

Paul M. Churchland

Dept. of Philosophy, UCSD

10/17/04



For those of us who were undergraduates in the 1960s, functionalism in the philosophy of mind was one of the triumphs of the new analytic philosophy. It was a breath of theoretical fresh air, a framework for conceptual clarity and computational rigor, and a shining manifesto for the possibility of artificial intelligence. Those who had been logical behaviorists rightly embraced it as the natural and more penetrating heir to their own deeply troubled views. Those who had been identity theorists embraced it as a more liberal but still agreeably robust form of scientific materialism. Those many who hoped to account for cognition in broadly computational terms found, in functionalism, a natural philosophical home. Even the dualists who refused to embrace it had to give grudging approval for its strictly anti-reductionist stance. It had something for everyone. Small wonder that it became, and has largely remained, the dominant position in the philosophy of mind, and, perhaps more importantly, in cognitive psychology and classical AI research as well.

Whether it still deserves that position – indeed, whether it ever did -- is the principal subject of this essay. The legacy of functionalism, now visible to everyone after forty years of philosophical and scientific research, has not been entirely positive. But let us postpone criticism for a moment, and remind ourselves of the central claims that captured so many imaginations.

I. THE CENTRAL CLAIMS OF CLASSICAL FUNCTIONALISM

- 1) What unites all cognitive creatures is not that they share the same computational mechanisms (their ‘hardware’). What unites them is that (plus or minus some individual defects or acquired special skills) they are all computing the same, or some part of the same, abstract << sensory input, prior state>, < motor output, subsequent state>> *function*.¹
- 2) The central job of Cognitive Psychology is to *identify* this abstract function that we are all (more or less) computing.
- 3) The central job of AI research is to create *novel physical realizations* of salient parts of, and ultimately all of, the abstract function we are all (more or less) computing.

¹ Just to remind, a function is a set of input/output pairs, such that for each possible input, there is assigned a unique output. Such sets can have infinitely many input/output pairs, and the relations between the inputs and outputs can display extraordinary levels of complexity. The characterization proposed in **1**) is thus in no sense demeaning to cognitive creatures. It requires only that the relevant function be computable, i.e., that the proper output for any given input can be recursively generated by a finite system, such as a brain, in a finite time.

- 4) Folk Psychology – our common-sense conception of the causal structure of cognitive activity -- already embodies a crude and partial representation of the function we are all (more or less) computing.
- 5) The reduction of Folk Psychology (indeed, any Psychology) to the neuroscience of human brains is twice impossible, because:
 - i) the relevant function is computable in a potentially infinite variety of ways, not just in the way that humans happen to do it, and
 - ii) such diverse computational procedures are in any case realizable in a potential infinity of distinct physical substrates, not just in the specifically human biological substrate.

Accordingly, to reduce the categories of Folk Psychology to the idiosyncratic procedures and mechanisms of specifically *human* brain activity would be to *exclude*, from the domain of genuine cognitive agents, the endless variety of other realizations of the characteristic function (see point **1**) that we are all computing. The kind-terms of Psychology must thus be functionally rather than naturalistically or reductively defined.

- 6) Empirical research into the microstructure and microactivities of human and animal brains is entirely legitimate (for certainly we do wish to know how the sought-after function is *realized* in our own idiosyncratic case). But it is a very poor research strategy for recovering the global function itself, whose structure will be more instructively revealed in the situated molar-level behavior of the entire creature

7) Points 5) and 6) jointly require us to respect and defend the *methodological autonomy* of Cognitive Psychology, relative to such lower-level sciences as brain anatomy, brain physiology, and biochemistry. Cognitive Psychology is picking up on its own laws at its own level of physical complexity.

Thus the familiar and collectively compelling elements of a highly influential philosophical position. Perhaps astonishingly, the position is decisively mistaken in all seven of the elements just listed. Or so, at least, I shall argue in what follows.

I. SOME UNEXPECTED LESSONS FROM NEUROBIOLOGY

The classical or ‘program-writing’ research tradition in AI was one highly promising expression of the Functionalist view just outlined. But by the early 1980s, that research program had hit the wall with an audible thud. Despite the development of CPUs with increasingly fabulous clock speeds (even desk-top machines now top 10^9 Hz), despite ever-expanding memory capacities (even desk-top machines now boast over 10^{10} bytes), despite blistering internal signal conduction velocities (close to the speed of light), and despite the continuing *a priori* assurance (grounded in the Church-Turing Thesis) that a Universal Turing Machine could, in principle, compute any computable function whatever, programmed computers in fact performed very poorly relative to their biological counterparts, at least on a wide variety of typical cognitive tasks.

The problem was not that there was any well-defined class of cognitive tasks that programmed digital computers proved utterly unable to even begin to

simulate. The problem was rather that equal increments of progress toward more realistic cognitive simulations proved to require the commitment of exponentially increasing resources in memory capacity, computational speed, and program complexity. Moreover, even when sufficient memory capacity was made available to cover all of the empirical contingencies that real cognition is prepared to encounter, a principled way of retrieving, from that vast store, all and only the *currently relevant* information proved entirely elusive. As the memories were made larger, the retrieval problem got worse. Accordingly, as actual cognitive performance approached the levels displayed by biological brains (and in many cases they did), the time taken for the machines to produce the desired performance expanded to ridiculous lengths. A programmed machine took minutes or hours to do what a biological brain could do in a fraction of a second.

At the time, this was deeply puzzling, because no process in the brain had a ‘clock frequency’ higher than perhaps 100 Hz, and because typical signal conduction velocities within the brain are no greater than a human bicycle rider: perhaps 10 m/sec. In the respects at issue, this puts the biological brain at an enormous disadvantage: $\approx 10^2$ Hz vs. $\approx 10^9$ Hz in the first dimension of performance, and ≈ 10 m/sec vs. $\approx 10^8$ m/sec in the second. All told then, the computer should have a computational speed advantage of roughly $10^7 \times 10^7 = 10^{14}$, or fourteen orders of magnitude. And yet, as we now say, shaking our heads in amazement, the presumptive tortoise (the biological brain) easily outruns the presumptive hare (the electronic digital computer), at least on a wide variety of typical cognitive tasks.

The explanation of the human brain's impressively high performance, despite the handicaps mentioned, is no longer a matter of any controversy. The brains of terrestrial creatures all deploy a computational strategy quite different from that deployed in a standard serial-processing, digital-coding electronic computer. That strategy allows them to do a clever end-run around their time-related handicaps. Specifically, the biological brain is a massively *parallel* piece of computational machinery: it performs trillions of individual computational transformations -- within the 10^{14} individual microscopic *synaptic connections* distributed throughout its volume -- *simultaneously* and *all at once*. And it can repeat such feats of computation at least ten and perhaps a hundred times per second. The presumptive deficit of fourteen orders of magnitude scouted earlier is thus made good in one fell swoop. And the brain is left with a modest computational advantage of its own concerning the number of basic computational operations performed per second: perhaps one or two orders of magnitude over current electronic machines.

Moreover, this massively parallel, distributed processing (or "PDP" as it has come to be called) provides a built-in solution to classical AI's chronic problem of how to access, in real time and from the totality of one's vast memory store, all and only the informational elements that are relevant to one's current computational problem. The fact is, the acquired strengths or 'weights' of the brain's 10^{14} synaptic connections collectively embody *all* of the acquired wisdom and acquired skills that the creature commands. (Learning, at least in its most basic form, *consists in* the progressive modification of those myriad connections.)

But those 100 trillion synaptic connections are also the brain's basic *computational* elements. Each time a large cadre of synaptic connections effects a transformation of an incoming representation into an output representation at the receiving population of neurons, *every synapse in that entire cadre* has a hand in shaping that computational transformation, and each makes its tiny contribution simultaneously with all of the others.

Accordingly, it is not just the brain's computational behavior that is massively parallel. Its *access to memory* is also a massively parallel affair. Indeed, these are no longer distinct processes, as they are in a digital computer with a classical von Neumann architecture. In the biological brain, to engage in any computational transformation simply *is* to deploy whatever knowledge it has accumulated. Accordingly, the classical Frame Problem² for Artificial Intelligence simply evaporates, as does the Inductive Logician's problem of the global sensitivity (to background knowledge) of any Abductive Inference,³ which is easily the most common form of inference that any creature ever performs.

These welcome developments concerning the general nature of information processing in humans and animals were humbling for the ambitions of classical AI not because those ambitions were revealed to be unachievable. On the contrary, artificial intelligence now looks more achievable than ever. Rather, these decisively illuminating developments were humbling because they were the result of empirical and theoretical research within two *lower-level* sciences,

² Dennett, "Cognitive Wheels: The Frame Problem in Artificial Intelligence," in Hookway, C., ed., *Minds, Machines, and Evolution* (Cambridge: Cambridge U.P., 1984).

Neuroanatomy and Neurophysiology, whose contributions to Cognitive Psychology and AI were widely and officially expected to be minimal at best, and procrustean at worst. (See again points 5), 6) and 7).) But those often-derided ‘engineering details’ turned out to be decisively relevant to understanding how a plodding biological brain could keep up with an electronic machine in the first place. And they proved equally decisive for understanding how the brain could routinely solve a vital *cognitive* problem – the real-time selective deployment of relevant information – that the programmed serial machines were quite unable to solve. Cognitive Psychology, it began to emerge, was not so ‘methodologically autonomous’ as the Functionalists had advertised.

III. FOLK PSYCHOLOGY AS A ROUGH TEMPLATE FOR OUR COGNITIVE PROFILE: SOME PROBLEMS

More generally, the perspective on cognition that emerges from neuroanatomy and neurophysiology holds out an entirely novel conception of the brain’s fundamental mode of *representation*. The proposed new unit of representation is the *pattern of activation-levels* across a large population of neurons (*not* the internal sentence in some ‘language of thought’). And the new perspective holds out a correlatively novel conception of the brain’s fundamental mode of *computation* as well. Specifically, the new unit of computation is the *transformation of one activation-pattern into a second activation-pattern* by

³ For a recent summary, see Fodor, “The Mind Doesn’t Work That Way” (Cambridge: The MIT Press, 2000). Also, Churchland, *Inner Spaces & Outer Spaces: The New Epistemology* (Cambridge: The MIT Press, forthcoming), ch. 2.

forcing it through the vast matrix of synaptic connections that one neuronal population projects to another population (*not* the manipulation of sentences according to ‘syntactic rules’). Since our own dearly-beloved Folk Psychology *shares* in classical AI’s linguaformal portrayal of human cognitive activity, the new vector-coding / vector-processing portrayal of our cognitive processes therefore casts the integrity of Folk Psychology into doubt as well, at least as an account of the *basic* structure of cognitive activity. Element **4**) of the preceding functionalist manifesto is therefore severely threatened, if not outright refuted, in addition to elements **6**) and **7**). Its warm familiarity and yeoman social service notwithstanding, Folk Psychology appears to embody no insight whatever into the *basic* forms of representation and computation deployed by typical cognitive creatures.

This is an outcome that we should have expected in any case, since we appear to be the *only* species of cognitive creature on the planet that is capable of deploying the syntactic structures characteristic of language. If *all* cognition deploys them as the basic mode of doing business, why are the other terrestrial creatures so universally unable to learn any significant command of those linguistic structures? And if the basic modes of cognition in those other creatures are therefore almost certain to be *nonlinguaformal* in character, then why should we acquiesce in the delusion that human cognition – alone on the planet – *is* linguaformal in its basic character? After all, the human brain differs only marginally, in its microanatomy, from other mammalian brains; we are all closely proximate twigs on the same branch of the Earth’s evolutionary tree. And the

vector-coding / vector-processing story of how terrestrial brains do business is no less compelling for the human brain than it is for the brain of any other species. We have here a gathering case that Folk Psychology is a modern cousin of an old friend: Ptolemaic Astronomy. It serves the purposes of rough prediction well enough, for an important but parochial range of phenomena. But it badly misrepresents what is really going on.⁴

IV. MULTIPLE REALIZATION: ON THE ALLEGED IMPOSSIBILITY OF AN INTERTHEORETIC REDUCTION FOR ANY MOLAR-LEVEL PSYCHOLOGY

Conceivably, the preceding estimate of Folk Psychology is too harsh. Perhaps its presumptive failure to mesh with the vector-coding / vector-processing story of brain activity reflects only the fact that Folk Psychology is a molar-level portrait of cognitive activity, a portrait that picks up on laws and categories at a level of description far above the details of neuroanatomy and neurophysiology, a portrait that should not be *expected* to reduce to any such lower level of scientific theory. As many will argue, that reductive demand should not be imposed on Folk Psychology – nor on any potential replacement Cognitive Psychology either (a replacement drawn, perhaps, from future molar-level research). For, it will be said, Psychology addresses law-like regularities at its own level of description. These regularities are no doubt implemented in the underlying ‘hardware’ of the

⁴ These skeptical themes go back a long way. See my “Eliminative Materialism and the Propositional Attitudes,” *J. Phil.* **78**, no. 2 (1981): 67-90. For even earlier doubts, see Feyerabend, P.K., “Materialism and the Mind-Body Problem,” *Review of Metaphysics*, **17** (1963): 49-66; and Rorty, R., “Mind-Body Identity, Privacy, and Categories,” *Review of Metaphysics*, **19** (1965): 24-54.

brain, but they need not be reducible to a theory of that hardware.⁵ For there are endlessly many different possible material substrates that would sustain the same profile of molar-level cognitive activity.

The claim that molar-level cognitive activities are multiply realizable is almost certainly correct. Much less certain, however, is the idea that multiple realizability counts against the possibility of an intertheoretic reduction of Folk Psychology, and against the reduction of any scientific successor Cognitive Psychology that is similarly concerned with intelligence at the molar level. The knee-jerk presumption has always been that any such reduction, to the underlying laws of any *one* of the many possible material substrates, would be hopelessly *chauvinistic*, in that it would automatically preclude the legitimate ascription, of the cognitive vocabulary being reduced, to entities composed of any of the many *other* possible material substrates. But this inference needs to be reexamined. It is, in fact, wildly fallacious.

What fuels the inference is the assumption that different material substrates – such as mammalian biology, invertebrate biology, extraterrestrial biology, semiconductor electronics, interferometric photonics, computational hydrology, and so on – will be governed by *different* families of physical laws. But this needn't be so. Let me illustrate with three salient and instructive examples.

⁵ Cf. J.A. Fodor, "The Special Sciences," **28**, *Synthese* (1974): 77-115.

Sound is a molar-level phenomenon. That is to say, it can be displayed only where there exists a large number of microscopic particles interacting in certain ways. And it, too, is a phenomenon that is multiply realized: in the Earth's highly peculiar atmosphere, in a gas of any molecular constitution, in a liquid of any molecular constitution, and in a solid of any molecular constitution. Sound propagates in any and all of these media. And yet sound is identical with, is smoothly reducible to, compression waves as propagated in any of these highly diverse media. For the underlying physical laws that bring the phenomenon of sound into the embrace of mechanical phenomena generally are *indifferent* to the peculiar molecules that make up the conducting medium, and to their collective status as a gas, liquid, or solid. What matters is that, collectively those particles form an *elastic medium* that allows energy to be transmitted over long distances while the elements of the transmitting medium merely oscillate back and forth a comparatively tiny distance in the direction of energy transmission. To put it bluntly, the very *same* laws of wave propagation in an elastic medium cover *all* of the diverse cases at issue. Idiosyncratic features such as the *velocity* of wave propagation may indeed depend upon the details of the conducting medium (such as the mass of its molecules, and whether they form a gas, liquid, or solid). But the various high-level laws of acoustics (such as $v = \lambda\omega$, and other laws concerning the reflective and refractive behaviors of sound generally) reduce to the very same mechanical laws in all of these diverse cases. A diversity of material substrates here does *not* entail diversity in the underlying laws that govern those diverse substrates. Accordingly, acoustics is not an 'autonomous

science,' devoted to finding laws and ontological categories at its 'own level of description.' It is but one chapter in the broader mechanics of elastic media.

Temperature, also, is a molar level phenomenon. And it, too, is a phenomenon that is multiply realized: in the Earth's atmosphere, or in any atmosphere, or indeed, in a gas of any molecular constitution whatever, either pure or mixed. For the temperature of a gas is identical with, is reducible to, the mean level of kinetic energy of the molecules that make up that gas. Here again, the underlying laws of motion (Newton's laws) that govern the behavior of, and the interactions of, the molecules involved are *the very same* for every kind of molecule that might be involved. Those laws are simply indifferent to the shape, or the mass, or the chemical makeup of whatever molecules happen to constitute the gas in question. Idiosyncratic details, such as the velocity of *dispersion* of an unconfined gas, will indeed depend on such details as molecular mass. But the laws of classical thermodynamics (such as the Ideal Gas Law, $PV = \mu RT$) reduce to the same set of underlying mechanical laws whatever the molecular makeup of the gas in question. Once again, a diversity of material substrates does *not* entail diversity in the underlying laws that govern those diverse substances.

Accordingly, classical thermodynamics is not an 'autonomous science,' devoted to finding laws and ontological categories at its 'own level of description.' Its reduction to statistical mechanics is a staple of undergraduate physics texts.

For a third example, a dipole magnetic field -- as instanced in the simple rectangular bar-magnet that one uses to pick up scattered thumb-tacks -- constitutes a molar-level phenomenon, but such dipole magnetic fields are

realizable in a variety of distinct metals and materials. Pure iron is the most familiar substrate, but sundry alloys (such as aluminum + nickel + cobalt) will also support such a field, as will certain metal/ceramic mixtures. Indeed, *any* substrate that somehow involves charged particles moving in mutually aligned circles (such as a tightly wound current-carrying coil of copper wire) will support a dipole magnetic field. For the simple laws that describe the shape and causal properties of such a field are all reducible to lower-level laws (Maxwell's Equations) that describe the induction of electric and magnetic fields by the motion of charged particles such as electrons. And those lower-level laws are, once again, indifferent to the details of whatever material substrate happens to sustain the circular motion of charged particles.

Once again, an open-ended diversity of sustaining substrates does not entail the irreducibility of the molar-level phenomenon therein sustained. And the historical pursuit of the various pre-Maxwellian theories of dipole magnetic fields (e.g., 'effluvium' theories) did not constitute an 'autonomous science,' forever safe from the reductive reach of new and more comprehensive theories. On the contrary, the work of Faraday and Maxwell brought those older theories into the welcoming embrace of the new, and much to the illumination of the former.

These examples can be multiplied. But even one such example illustrates the hazards of inferring the irreducibility of some molar-level phenomenon from the premise of its multiple realizability, even when that premise happens to be *true*. For diverse material substrates may still be governed, all of them, by some of the very *same* low-level physical laws, laws quite capable of explaining the

molar-level behaviors *shared* by all of those diverse substrates. The classical argument for element 5) of the functionalist manifesto is plainly flawed, for it naively ignores these obvious sorts of reductive possibilities.

Let us push this line of thought a little farther. For it now begins to appear that, beyond merely providing decisive counter-examples to what Functionalism presents as a robustly *nonreductive* pattern, the three examples just cited also provide an *alternative* pattern – a pattern whereby molar-level theories that record genuine high-level regularities across diverse material substrates will *typically* find a successful reduction to some underlying and highly general physical laws, laws that are simply blind to the idiosyncratic and irrelevant differences that happen to distinguish the several substrates. We have just seen this happen in three unproblematic cases. And there are, I repeat, many more.⁶

⁶ For a fourth example, consider the shared molar-level thermodynamic profile of living organisms across a wide variety of biochemical substrates. The underlying laws of nonequilibrium thermodynamics are once again blind to the peculiar chemical makeup of such diverse substrates. For a fifth example, consider Kepler's three laws of planetary motion, valid for planets of highly diverse material constitution. All three of those laws are reducible to Newton's particle mechanics plus his Universal Law of Gravitation. For a sixth example, consider the science of aerodynamics: the theory of creatures or machines that are capable of flight. Multiple realizability is an obvious feature of this domain: think of seagulls, hummingbirds, bats, dragonflies, wooden airplanes, metal airplanes, helicopters, and so on. And yet their shared molar behavior is ultimately owed to the fact that they all contrive to accelerate ambient air more-or-less continuously downwards, which activity yields, by Newton's (substrate-neutral) third law, a reactive upwards force that is more-or-less continuously equal to the task of keeping them aloft. For a seventh example, consider the closely similar chemical and electrical behaviors of the distinct elements within a given chemical 'family,' those that constitute one vertical column of the Periodic Table (e.g., the metals, or, the noble gases). Here the shared molar-level chemical regularities, across a given family of elements, are explained in terms of shared valence-electron-shell structures across the distinct types of atoms within that family. For an eighth example, one rather closer to the case of cognitive creatures, consider the molar behavior of any radio, TV, or music player. Despite the great variety of metal and semiconductor substrates that will instantiate the required circuits for signal detection, amplification, and presentation, the behavior of all such devices is reducible to the same set of electro-dynamical laws concerning resistances, capacitances, and inductances, laws blind to the material diversity of the substrates that a given manufacturer may choose to employ. As we see from such examples, this general *reductive* pattern, across substrate diversity, is quite robust. For an illuminating discussion, see Strevens, M., *Bigger than*

On this alternative logical and historical pattern, legitimate molar-level theories that comprehend genuine natural kinds will thus be positively *expected* to find some such intertheoretic reduction. For if they eventually prove *not* to be thus reducible, we will have to reconsider the initial presumption that the molar-level theory really does embrace genuine high-level natural kinds governed by genuine high-level explanatory laws. The ‘unitary’ account that the molar theory seemed to provide, across the diverse substrates, might then have to be judged an accidental or a false unity. And its supposedly law-like generalizations will thus turn out to be accidental generalizations of some sort, generalizations that are empty of real explanatory and predictive power. Accordingly, if we expect our beloved Folk Psychology, or *any* Psychology, to provide an accurate, natural-kind-embracing, genuinely nomological and explanatory account of the molar-level cognitive operations and behavior of humans, other mammals, human-like aliens, and human-like artificial automata, then we had better *hope* that there exist highly general underlying laws – laws blind to the material differences between all of these diverse creatures – which serve collectively to explain, and thus to reduce, the categories and laws of Psychology.

Let us finally confront the most important question here at issue. Just what are the chances of finding some substrate-neutral underlying laws – laws with a suitably broad explanatory reach – for *Psychology* in particular? That is to

Chaos (Cambridge: Harvard U.P., 2003), especially ch. 5, “Implications for the Philosophy of the Higher-Level Sciences.”

say, what are the chances that the case of Psychology will turn out to be an instance of the alternative and overtly *reductive* pattern of development explored in the preceding pages, and in the examples of footnote 7? Well, they are certainly not zero. For there are at least two low-level theories that have sufficient generality to embrace all of the diverse material realizations of cognitive activity listed in the preceding paragraph, and that also hold promise for explaining at least some of the activities comprehended by Psychology. Let us take a look at them.

V. SOME REDUCTIVE POSSIBILITIES FOR MOLAR-LEVEL PSYCHOLOGY

The first possible framework equal to the task of comprehending all of the diverse material realizations envisioned for cognitive systems is Nonequilibrium Thermodynamics. Distinct from the more familiar (near-equilibrium) statistical thermodynamics discussed earlier, this is the general framework for describing the laws of energy and information flow in partially-closed physical systems that are, and remain, very far away from energetic equilibrium. This is the framework, still in its developmental infancy, that already unites and illuminates all *biological* phenomena, whatever their physical constitution. The basic idea, first outlined half a century ago by the physicist, Erwin Schrodinger,⁷ is that any living organism is a highly improbable physical structure whose natural behavior – if it is located in a suitable flow of ambient energy -- serves to exploit whatever structure it already contains so as to produce additional physical structure. It

grows, or it repairs itself, or it reproduces. Such an interest-bearing investment⁸ is possible only when the system is situated so as to exploit an energy flux that begins with energy from a very low-entropy state⁹, energy that is then progressively dissipated into energy at a much higher entropy state. The living physical system ‘steals’ some of the initial low-entropy energy as that energy courses through it, and it then incorporates that energy in the form of additional (and improbable) physical structure. The low-entropy energy source for our terrestrial environment is ultimately the Sun, radiating at a black-body temperature of roughly 4000°K (i.e. at rather short wavelengths). And the ultimate high-entropy energy sink is the surrounding background of empty space, radiating at a black-body temperature of about 3°K (i.e., at very long wavelengths). In between lies the biosphere at a temperature around 293°K. Without such a concentrated or low-entropy energy source ‘above’ us, and such a dissipated high-entropy energy sink ‘below’ us, nothing alive could hope to remain alive. Indeed, nothing of any biological interest – that is, no extremely improbable physical structures with complex metabolic pathways -- could ever have evolved in the first place.

Those that have evolved are thus instant testament to the existence of such a complexity-inducing ambient energetic waterfall – a constant flow from the Sun, through us, and into the cold abyss beyond. Moreover, any individual of any

⁷ E. Schrodinger, “What is Life?” (Cambridge U.P., 1944).

⁸ Note well the economic metaphor here deployed. Its aptness will come up again shortly.

⁹ Entropy is a measure of how chaotically *scattered or dissipated* an amount of energy happens to be, a measure of how *unavailable* it is to do any concerted work. By contrast, a low-entropy state implies that the relevant energy is highly ‘concentrated’ and available to do work.

species also embodies, in its typical structural details, extraordinary amounts of information about the peculiar environmental niche in which it thrives. For no individual could be expected to *have* the specific physical structure it has *unless* the environment in which it thrives has a comparably specific physical and dynamical profile. For, once again, eons of evolutionary pressures have made the former exquisitely ‘tuned’ to the latter, functionally and metabolically speaking. If you want an indirect but highly informative window onto the chemical and biological dynamics of a forest pond, examine the frog who lives there.

I provide the reader with this brief sketch, of how nonequilibrium thermodynamics embraces biological phenomena generally, not just to validate the claim made in footnote 7, but to prepare the ground for the following suggestion. Cognitive phenomena are just an additional instance or iteration of the dynamical profile already outlined. Specifically, a creature that *learns* about the world is a creature that exploits the low-entropy internal structure or information that it already possesses, in such a fashion that, if the creature is placed in an environment with an information-rich energy flow, it comes to embody *additional* information-bearing structure about its environment, typically in the form of a progressively re-wired brain.

The relevant energy flow here begins with the low-entropy states of incoming *sensory* signals (light, sound, pressure, taste, smell, whatever), signals that contain detailed information about the immediate physical environment. And it ends with the dissipation of that original low-entropy energy, after the brain’s cognitive processing is done, as high-entropy heat radiated away by the body at

long wavelengths in the infrared, its original information ‘lost’, or rather, left behind in the brain. The active cognitive system ‘steals’ some of the low-entropy energy that its sensory organs provide, and incorporates it as additional information-bearing structure. My biological body at age six days will embody a great detail of general information about my natural environment, as we noted two paragraphs ago, concerning the frog. But my brain, at six years (or at six decades), embodies an additional wealth of information, information that the human genome was, and is, far too small to bequeath to me. I have to acquire that wealth of information post-natally. And I do. And so does a mouse. And so does any cognitive creature. For that is what a cognitive creature is: an ‘extra-somatic information multiplier’. Unlike a typical Heat Engine (e.g., a steam engine, an automobile engine), which exploits an entropy-increasing energy flow to produce macroscopic *motion*, the brain exploits such a flow to produce neuronally and synaptically-embodied *information*. We are, in fact, *Epistemic Engines*, not just figuratively, but literally.

This naturalistic portrait, note well, makes no reference to the variety of material substrates that might sustain such an energetic and informational economy. Many different substrates are presumably possible. What makes them all cognitive creatures – part of it, anyway -- is their shared thermodynamic and information-multiplying profile.

I briefly floated this possible construal of cognitive creatures some twenty-five years ago, in the closing paragraphs of my 1979 book,¹⁰ and again, in slightly more detail, in a 1982 paper.¹¹ Those accounts of cognition, and that of the preceding two paragraphs, may well be dismissed as mere hand-waving speculation, unless we can provide an account of *how* brains actually process, and incorporate into their internal structure, ambient information.

In the salient case of *biological* metabolisms, we do indeed possess such a non-handwaving account. We know how DNA embodies information. We know how that information is read out by RNA in order to synthesize various protein molecules. We know how those protein molecules catalyze certain metabolic reactions, and sequences of such reactions. We know how those reaction-chains create new biological molecules that form additional biological structures. We know how those structures collectively steer ambient energy and materials along paths that sustain and amplify the organism at issue. The nonequilibrium thermodynamical portrait of living things is therefore not just a philosophical guess. It is a highly general reductive framework that brings real illumination to biological processes, across a wide diversity of chemical substrates.

Twenty-five years ago, I must own, the nonequilibrium thermodynamical portrait of *cognitive* activity *was* a merely philosophical guess – a hesitant extrapolation from the thermodynamic portrait of living things just explored. For we then lacked any corresponding account of how brains actually process and

¹⁰ *Scientific Realism and the Plasticity of Mind* (Cambridge: Cambridge University Press, 1979).

¹¹ “Is *Thinker* a Natural Kind?”, *Dialogue* **21**, no. 2 (1982).

eventually store new information. In the intervening years, however, an account of exactly those activities has pieced itself together, and has become the focus of a great deal of research, both experimental and theoretical. That account posits fleeting *activation vectors* across proprietary populations of neurons as the basic mode of ephemeral representation. It posits *vector-to-vector transformations*, at the hands of intervening matrices of synaptic connections, as the basic mode of computation over those representations. It posits specific *configurations of weighted synaptic connections* as the basic mode of general or background knowledge. And it posits ongoing *adjustments in the values* of those weighted synaptic connections as the most basic mode of learning.¹²

It is vital to appreciate that the structural and dynamical portrait just painted – of vector coding and vector processing via large matrices with plastic coefficients – is once again a portrait that can be realized in a wide variety of material substrates: in mammalian brains, in octopus brains, in extraterrestrial brains, in electronic chips, in optical systems, and so forth. For the mathematical laws of vector/matrix processing, and the information-dependent, experience-driven adjustments of the individual coefficients of the transforming matrix, are all indifferent to the physical medium that displays those fleeting vectors and embodies those comparatively enduring matrices. Some idiosyncratic details, such as cognitive reaction times, will indeed be sensitive to the implementational

¹² For accessible, entry-level accounts that will provide doors onto the wider literature, see Churchland, P.S. and Sejnowski, T.J., *The Computational Brain* (Cambridge: The MIT Press, 1987); Churchland, P.M., *The Engine of Reason, The Seat of The Soul* (Cambridge: The MIT Press, 1995); Churchland, P.M., *Inner Spaces and Outer Spaces: The New Epistemology* (in preparation).

facts, such as the speed of vectorial conduction between distinct populations of active units. (As noted earlier, an electric current in a copper wire propagates much faster than a spike train in an axon.) But the profile of molar-level activity will be importantly similar across all of these diverse substrates.

Once again we are contemplating a low-level explanation, in terms of general or abstract underlying natural laws, of a roughly constant profile of molar-level activity, activity that can be displayed across a considerable variety of material substrates. But this time the explanatory target is the *cognitive* activities of creatures like ourselves. We are no longer pressing a mere *a priori* possibility on our functionalist friends. We are confronting a pair of real theories (the vector/matrix account of brain structure and function, and the nonequilibrium thermodynamical account of brain growth and learning), theories that hold some nontrivial promise of providing systematic reductive explanations of cognitive phenomena in particular, despite their presumably diverse realizations. To put it bluntly, we are confronting exactly what classical functionalism said was not to be had, nor even to be sought.

These new developments, especially the vector/matrix story, have already given us a much deeper understanding of what the brain does, and of how it manages to do it. We now understand, for example, how the activation space of a large population of neurons can come to embody a structured system of categorical prototypes – that is, a meaningful *conceptual framework*. We now understand how those prototype-points in such a background neuronal activation space can be selectively activated by sensory inputs. That is, we understand how

a brain can *interpret* its sensory experience in terms of its acquired conceptual framework. We now understand how prototypical motor behaviors can be represented as prototypical activation-trajectories in motor-neuron activation space. That is, we have some understanding of how complex *motor skills* are embodied. And we know how such unfolding trajectories can actually *generate* the relevant motor behaviors in the body's limb and muscle systems. In sum, we can now see cognitive activity as we have never seen it before. Whether we are seeing it correctly, only time will tell. But a fertile vision is already being explored. What this means is that the celebrated element 5) of the functionalist manifesto is not just naively argued. In fact, it is almost certainly false.¹³

VI. WHAT DOES *NOT* UNITE THE CLASS OF COGNITIVE CREATURES

On the vector/matrix story explored above, what carries the burden of any creature's acquired background knowledge, of the world's general and enduring structure, is the specific configuration of the billions or trillions of synaptic connections that variously intervene between the brain's many distinct neuronal coding populations. It is these variously weighted excitatory or inhibitory

¹³ Allow me a closing remark on Fodor's 1974 parade case of a molar-level natural science for which reductive aspirations are supposed to be clearly foolish, namely Economics. The supporting argument then appealed to the multiple realizability of *currency* systems – such as shell currency, coin currency, paper currency, electronic currency, and so on. We can all agree that Economics is not going to be reducible to the chemistry of wood-fiber, or to the physics of copper and gold. But all of this is now visibly beside the point. For we can now appreciate that Economics is the study of *the metabolisms of superorganisms*, a phenomenon that once again falls firmly within the province of nonequilibrium thermodynamics, a science whose laws are blind to such implementational details. A national economy, after all, embodies a flow of both energy and materials: it creates real physical and organizational structures, and it dissipates vast amounts of (initially low-entropy) energy in the process. It is too soon to *insist* that Economics will indeed

elements that constitute the brain's principal memory store, and also its principal means of computation. One and the same system simultaneously serves both functions

No two people, however, display the same configuration of synaptic connections and synaptic weights. Each human brain boasts roughly 10^{14} synaptic connections, the overwhelming majority of which are established post-natally in response to a lived experience that is unique to each individual. Since we experience a common world that does display enduring features, each of us ends up with a family of sculpted activation-spaces whose structure is *similar* to the structure of other people's activation spaces, at least if they are members of the same culture. But genuine identity is too much to ask for. We may all agree that the vector/matrix system found in each individual is computing a function of some fabulously complex sort. But no two people on the planet will be computing exactly the same function, for no two people share the same matrix of synaptic connections.

Very well, but surely they will be computing *similar* functions? Indeed so, if they happen to be peas from the same cultural pod.¹⁴ But what wants emphasizing here is the real functional *diversity* displayed by individuals at different stages of their lives, in different cultures across the planet, and at different points in our very long and cognitively diverse human history. This

find such an explanatory reduction. But neither can Fodor justly insist that it won't. The presumption in favor of his principal nonreductive example has just evaporated.

¹⁴ How this similarity can be achieved, despite our synaptic diversity, is detailed in Churchland, P.M., "Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered," *this journal*, **XCIV**, no. 1 (Jan., 1998): 5-32.

diversity in the functions that brains are computing becomes more striking still when we expand our consideration to include cognitive activity in the nervous systems of other terrestrial creatures such as chimpanzees, cats, mice, finches, crocodiles, crabs, octopi, fish, and spiders. Clearly such diverse creatures are not all computing the same function, nor even remotely similar functions. And yet, we are all cognitive creatures.

What is it, then, that unites us? Ironically, it appears to be the abstract form of our *hardware* that unites us! We are, all of us on the preceding list, massively parallel vector-processors whose ever-active vector-transforming matrices (our trillions of synaptic connections) are slowly updated or instructed by a procedure that filters information from a low-entropy flux of energy from our sensory peripheries. This computational arrangement has prodigious advantages over the serial architecture deployed in classical (von Neumann) computers – in its speed of computation, in its graceful tolerance of scattered component failures, and in its fast retrieval of relevant information. This alternative computational template is sufficiently virtuous to make it a likely evolutionary choice on any planet that develops life, not just on Earth, and to make it a compelling technological choice for any future attempts at constructing artificial intelligence as well.

Accordingly, element 1) of the original functionalist manifesto is almost certainly a mistake – indeed, a monumental mistake. The cognitive creatures on this planet are computing a bewildering *variety* of very different functions, but they are all using fundamentally similar computational ‘hardwares’ in order to do

it. On this fundamental point, classical functionalism had things exactly backwards.

Element 2) must therefore be rejected as well. If our alternative portrait of cognition is even roughly correct, the central job of Cognitive Psychology is to explore *how* it is that terrestrial brains are able compute the extraordinary *variety* of functions displayed in diverse species of cognitive creature. This must be an empirical undertaking, one sensitive to the idiosyncrasies of nonhuman nervous systems. Accordingly, element 3) must be rethought along the same lines. The central job of AI research is not just to explore the construction of artificial vector-processing systems that compute the same function that some species of animal is already computing. A central part of its job will be to explore instead the pregnant potential of such artificial systems for computing functions – for pursuing cognitive activities – that no terrestrial creature has yet pursued or ever will pursue. Large-scale electronic realizations of our vector/matrix-style of computational resources will explore entirely new horizons for information processing and world-representation, and, being electronic, they will do it roughly a million times faster than biological creatures can ever hope to do it (because the speed of signal conduction in a copper wire is close to the speed of light). The enterprise of Artificial Intelligence thus has a dazzling future, but not because classical functionalism launched it in the right direction.

Indeed, it launched the enterprise in a most unfortunate direction. Element 6) provided a twisted rationale for mostly *ignoring* the empirical or experimental neurosciences, and for ignoring the early theoretical work that

attempted to model the activities of large numbers of interconnected neurons. Worse still, element 7) celebrated this deliberate disconnection with an ill-conceived positive portrait of Cognitive Psychology and Artificial Intelligence as ‘methodologically autonomous sciences.’ In retrospect, this was unwise, despite the genuinely clever contributions of a great many gifted researchers. For it served to insulate the relevant research from exactly the empirical information that promised the most interesting and authoritative constraints on whatever models were put up for evaluation. The result was almost half a century of misdirected research.

VII. CONCLUSIONS

The seven elements of our opening functionalist manifesto, all seven of them, appear to be false – not just inadequately argued for, but outright false. Fortunately, we now have an alternative positive portrait of cognitive activity in place, one capable of steering systematic research on the nature of cognition, research that points in directions interestingly different from the directions that dominated the last half of the Twentieth Century. That alternative portrait is no longer imprisoned by the linguaformal representational paradigm provided by Folk Psychology, nor is it centered on the computational paradigm provided by the digital-coding, serial-processing, program-running machines of the still-standard von Neumann configuration. Instead, it draws its opening inspirations, in both cases, from the deeply instructive empirical example of terrestrial nervous systems.

In closing, it is worth pointing out that two prominent background assumptions of the functionalist program have not been denied in the preceding critique. The first assumption is that cognitive creatures are indeed engaged in computing complex functions of some sort or other. And the second is that these computational activities, whatever they are, can be realized in a diversity of physical substrates. These assumptions are presumably as true, and as important, as they ever were. But in the present intellectual environment, those same two assumptions now pull our imaginations in entirely new directions. The first assumption motivates the brain-centered research program known as Computational Neurobiology. And the second assumption motivates the development of alternative physical realizations (presumably electronic or photonic) -- not of our 'software' (strictly speaking, we don't *have* any!), but -- of the massively parallel, vector-processing structure of our biological *hardware*. Let us hope that this second wave of research will be more revealing, and less self-blinkered, than the functionalist-inspired wave that preceded it.