

# Of Mind and Meaning

**Robert P. Doede**

**ABSTRACT** Key Words: cognitive science, naturalization, original semantics, personal agent, event-causation.

*This paper examine's Jerry Fodor's attempt to naturalize the human mind by encompassing it within a new mechanistic ontology. It then explores Polanyi's view of mind's embodiment and meaning's emergence in an effort to uncover some fundamental incoherencies in Fodor's naturalization project.*

The modern digital computer has revolutionized the way we think about thinking and mind's ontology. The computer (or so some claim) is an example of an information processing system that might (given the appropriate program) share our psychology, without also sharing our physical organization (Fodor 1981:9). Jerry Fodor, a pioneer in cognitive science, is convinced that the modern digital computer holds the secret to naturalizing the human mind, i.e., rendering it explicable in terms of natural science.

In an attempt to extend the 16th century metaphor of the world-machine to encompass human cognition, Fodor seeks to exploit the computer as an existence proof, demonstrating that cognition is reproducible by machines: machines that manipulate symbols according to rules sensitive only to the symbols' physical structure. Fodor, however, gives the machine metaphor a new twist. No longer is the paradigmatic machine a clock, a physical contraption subject to physical laws. Rather, Fodor appeals to what is known as a "virtual machine," an abstract device whose structure consists of abstract rules that operate on explicit symbols.

Fodor deploys the concepts developed in the modern synthesis of linguistics, proof theory, information theory, artificial intelligence, and neuropsychology, known as Cognitive Science (henceforth, CS), to yield a functional and naturalized (i.e., non-anthropomorphic) analysis of intelligence: one that is independent of physics and yet wholly mechanistic in intent. But, as I hope to demonstrate in the following pages, Fodor's CS nonetheless rests upon an anthropomorphism, however more rarefied it may be: instead of ascribing human properties directly to nature, he first ascribes the properties of human artifacts, viz., virtual machines, to nature (i.e., our neurophysiology), and then appeals to these properties in his effort to naturalize the mind.

In the pages to follow, I not only explore Fodor's project of naturalizing the human mind, I also develop (what I take to be) Michael Polanyi's view of mind's embodiment. I believe Polanyi's views illuminate some of the fundamental inadequacies not only in Fodor's program, but in any account of the human mind pitched at a sub-personal level, i.e., a level of analysis that goes beneath the conscious intentions, actions, and avowals of the personal agent and purports to explain these in terms of processes and events to which the personal agent lacks both intentional and conscious access.

## Fodor on Human Intelligence

Quite simply, Fodor believes that human intelligence is a symbol-driven, rule-governed, physical phenomenon (Fodor 1981: 23-24). That is, as much as H<sub>2</sub>O is the scientific essence of water, so too the essence of intelligence turns out to be manipulation of symbols according to rules. Fodor contends that “the only psychological models of cognitive processes that seem even remotely plausible represent such processes as computational” (Fodor 1981: 27). He agrees with his friend and fellow cognitivist, Zeno Pylyshyn, that cognition and computation are “species of the same genus” (Pylyshyn 1984: xiii), i.e., both are rule-governed operations on a symbolism. Let’s take a moment and unpack the idea of “computational process.” What exactly is it?

The short answer to this question is “anything that is Turing-machine computable.” But, again, what does it mean to be “Turing-machine computable”? Suffice it to say that any question whose answer may be gained through an algorithm is Turing-machine computable. An algorithm is a rule that is free of ambiguities, vagueness, and approximation, and that, in a finite number of steps, answers a question. Algorithms are rules whose applications to (pre-specified) variables result in definite conclusions, “yes” or “no” answers.

Propositional logic is a good example of a computational process. It has symbols that function as variables, and rules specifying the operations that might be performed on them. There are three things to notice about propositional logic that are characteristic of all computational processes (here see Haugeland, 1985: 99ff):

- (1) It is a formal system: a formal system has syntax alone; it is bereft of semantics. Its symbols and rules constitute a strictly notional world, delineating logical possibility, and thus bearing no relations to the contingencies of the empirical world. That is, a formal system isn’t about anything until its symbolic atoms are assigned interpretations: e.g., “p” stands for any atomic proposition in any natural language.
- (2) It is a perfectly definite system: its primitive data-structures have truth values that are wholly explicit, without ambiguities, vagueness, or approximations (either “T” or “F” is assigned to its primitives).
- (3) It is a finitely checkable system: all of its permissible transformations may be proven valid by reference to a finite number of rules (algorithms).

All computational processes are *digital* in nature: a digital system is a system whose every state is discrete (not continuous or partial)—a system that moves from state to state by sudden jumps from one definite state to another. All digital systems can, therefore, be automated, i.e., the algorithms describing their state transitions can be *mechanized*. Here is the link to Turing machines: if the mental processes responsible for human intelligence are really computational processes, then these mental processes will, like all computational processes, have purely syntactical necessary and sufficient conditions, and thus be wholly susceptible to mechanistic *cum* naturalistic explanation—human intelligence will be mechanistically generable. All one has to do is discover the program that underlies the intelligence and run it on a universal Turing machine.

A *universal* Turing machine is a *virtual machine*: a machine made of rules not wires, an abstract specification of the functional structure of a very simple computing device that can functionally imitate any computational system. The idea of a universal Turing machine helped entrench the now common distinction between a machine's software and its hardware. It showed how a very simple hardware could embody an unbounded number of softwares; how a single gadget could become a universal imitator of any and every instance of formal information processing; how very complex information processing reduces to some very simple mechanizable operations; how the brain (neuroware) could think (run) an unbounded number of different thoughts (programs).

A home computer is a concrete example of a universal Turing machine that, in some wholly *natural* fashion, harnesses the causal structure of its hardware to realize the computational operations represented by the software run on it. So it would seem that if Fodor and the cognitive scientists are right, and cognition and computation *really* are "species of the same genus," then Turing machines reveal how we might go about naturalizing cognition by conceiving it to be just a biological instance of the software-hardware interface found in one's home computer. This, in fact, is precisely the line Fodor wishes to follow.

## The Computer's Turing Machine: The Machine Code

Exactly how does software harness a computer's hardware? This is a long and complicated story. Fortunately, to understand Fodor's computational theory of mind, we need only concern ourselves with a minimum of technical notions.

First, let's consider the notion of hardware. Hardware is the "nuts and bolts." On a conventional computer, the hardware is the physical stuff whose behavior is dictated by physical laws; it is subject to the laws of gravitation, Ohm's law, etc. The hardware usually consists (among other things) of a central processor composed of silicon microprocessors and all kinds of electronic gadgetry that is engineered (hardwired) so that the on's and off's of its microswitches electronically mimic the binary states (T's and F's) of three Boolean logical operations, AND, OR, and NOT.

Now add the notion of machine language (henceforth, ML). A computer's ML is the abstract binary specification of the computer's most basic functional structure, the way in which its hardwired logic gates (microswitches) are linked. This is why ML may be viewed either as the top of the hardware hierarchy, or as the bottom of the software hierarchy. As a built-in code of primitive symbols and rules, it enables the hardware to receive instructions (each instruction written in machine code elicits a unique operation within the mechanism) from programmers (via assembly language), and thus serves as a universal Turing machine, enabling the computer to imitate *any* computable function. In the end, *the ML determines the state of the computer*; every instruction a programmer gives to the computer must ultimately be encoded into the monotonous strings of "0's" and "1's" that define its ML and dictate its state transitions and output. The only things the machine code can be used to talk about are the states and processes of the computer itself (it is the computer's language of self-description). So here is a picture of a machine whose state transitions aren't determined by (although dependent upon) physical laws: the computer moves from discrete state to discrete state according to the dictates of the electromagnetic syntax of the ML that its circuitry has been engineered to embody.

We are ready now to consider Fodor's human equivalent of ML, what he calls "the language of thought" or "Mentalese." The language of thought, on Fodor's view, is the human universal Turing machine whose primitive neurophysical symbols and rules drive human intelligence (Fodor and Pylyshyn 1988: 13).

## **The Human Turing Machine: The Language of Thought**

Fodor considers how thought could be *about* something else, how one bit of the world (a brain state) could be *about* another bit of the world (a state of affairs). He's convinced that thought, like language, must have a medium of representation, and that this medium of representation, like language, must also have a rule-governed structure (a grammar). Without a language-like medium of representation, he argues, we would not have the slightest idea how our thoughts could *mean* something or be true or false. Moreover, he contends that, if this medium of representation didn't have a rule-governed structure, we would have no idea how thoughts could combine and decompose leading to valid inferences. He is certain that we can think *about* things, make rational choices, perceive the world, or learn concepts, only because we are born (innately endowed) with a hard-wired medium of structured representation--"the language of thought" (henceforth, LOT).<sup>1</sup> In short, if we can think, perceive, or make decisions (all inference-driven behaviors, according to Fodor (Fodor and Pylyshyn 1988: 30), we *must* have an internal propositional code in which we formulate hypotheses about the world, build models of the world, and test theories about the world. That is, the etiology (causal antecedents) of intelligent behavior is a series of computational transformations of propositionally-structured information embodied in human hardware, the electrochemical transformations of LOT's symbols is taking place in the central nervous system (CNS) (Fodor 1975: 52).

Fodor takes a child's learning of a natural language as an example of concept learning: he believes all concept learning is accomplished by formulating hypotheses and then testing them in the representational medium of LOT (Fodor 1975: 58). One cannot learn a language, according to Fodor, unless one already knows one (Fodor 1975: 65). And the language we all already know prior to learning our natural language is the LOT. By virtue of this innate endowment, this private but universal Turing machine, a child learns its mother tongue:

Learning a language . . . involves learning what the predicates of the language mean. Learning what the predicates of the language mean involves learning a determination of the extension of these predicates. Learning a determination of the extension of the predicates involves learning that they fall under certain rules (i.e., truth rules). But one cannot learn that P falls under R unless one has a language in which P and R can be represented. So one cannot learn a first language unless one already has a system capable of representing the predicates in that language *and their extensions*. And, on pain of circularity, that system cannot be the language that is being learned. But first languages are learned. Hence, at least some cognitive operations are carried out in languages other than natural languages (Fodor 1975: 63-64).

Fodor appeals to the relation of ML and programming language to give some empirical plausibility to his rather incredible view of first language learning. In a computer, the ML is a universal Turing machine that functions as a metalanguage encoding the instructions of programming language (object language). So too, claims Fodor, the predicates of the natural language that people learn are encoded in the LOT in which computations are performed that are token-identical to human thought (Fodor 1975: 66). So learning a first language is a matter of both compiling (encoding) a natural language into one's LOT and then using one's LOT to formulate hypotheses about the extensions

of its predicates--quite a complicated task for a youngster to perform. Fodor, however, isn't too concerned about how a youngster, wholly ignorant of formal grammars and hypothesis-testing methodology, could perform the rather sophisticated grammatical and inferential tasks Fodor's theory requires of him or her. Fodor simply maintains that all this fancy foot work is the unconscious and, therefore, subpersonal doings of the youngster's built-in neurophysiologically-embodied LOT (Fodor 1975: 52).<sup>2</sup>

He admits that this position may appear scandalous, but refuses to take it as a *reductio* because no serious alternative to his theory of first language learning has ever been proposed (Fodor 1975: 82). He is quite willing to "bite the bullet" here, so long as he can "eat his cake" later (Fodor 1975: 52). I propose, in a later section, to point the way to a Polanyian alternative to his computational account. But before I turn to Polanyi, I shall outline Fodor's account of how the syntactical scaffolding of the LOT acquires semantic content.

### **Fodor on How the LOT Acquires a Semantic Dimension**

The division of hardware/software intrinsic to the cognitivist commitment to the computer model yields two networks of relations, the causal and the inferential. Fodor's idea is that if we can establish even a partial isomorphism between the two, we have a way of understanding the inferential regularities that underpin human intelligence, make a science of cognition possible; i.e., why, for instance, we can know that Bill believes (if only tacitly) that "the sky is clear" if we also know that Bill believes both that "either it is raining outside or the sky is clear," and that "it is *not* raining out." The problem is finding a *naturalistic* means of linking the causal mechanisms of human neuroware (brain and CNS) to the inferential regularities of human software (psychology), regularities that are sensitive to semantic properties. Fodor, relying on his argument that cognition requires a language-like medium of representation, contends that "computers show us how to connect semantical with causal properties *for symbols*" (Fodor 1985: 93).

On Fodor's view, the causal properties of a symbol in the LOT (i.e., the brain's neurophysiologically embodied equivalent to binary digit in a computer's ML) get connected to the symbol's semantic properties in a mind's thought life (i.e., the mind's equivalent to an instruction in a computer's program) via the symbol's syntax. Fodor explains:

The syntax of a symbol is one of its second-order physical properties. To a first approximation, we can think of its syntactic structure as an abstract feature of its (geometric or acoustic) shape. Because, to all intents and purposes, syntax reduces to *shape*, and because a symbol is a potential determinant of its causal role, it is fairly easy to see how there could be environments in which the causal role of a symbol correlates with its syntax. It's easy, that is to say, to imagine symbol tokens interacting causally *in virtue of* their syntactic structure. The syntax of a symbol might determine the causes and effects of its tokenings in much the way that the geometry of a key determines which locks it will open (Fodor 1985: 93).

So now all Fodor has to do is find a way of construing the syntactical relations between symbols so that they mirror the semantical relations holding between symbols. And his means of doing this are ready-to-hand: proof-theory, seen from a very great distance, does exactly this (Fodor 1985: 93). Proof-theory shows that, for formal languages, "the semantic relation that holds between two symbols when the proposition expressed by the one is implied by the proposition expressed by the other can be mimicked by syntactic relations in virtue of which one of the symbols is

derivable from the other” (Fodor 1985: 93). Thus

It would not be unreasonable to describe Classical Cognitive Science as an extended attempt to apply the methods of proof theory to the modeling of thought. . . . Classical theory construction rests on the hope that syntactic analogues can be constructed for nondemonstrative inferences (or informal, commonsense reasoning) in something like the way that proof theory has provided syntactic analogues for validity (Fodor and Pylyshyn 1988: 29-30).

Armed with the notion of a Turing machine, all one has to do is imagine a machine whose state transitions are dictated solely by the syntactic properties of its symbols, and whose operations on its symbols are entirely alternations of their shapes. If this machine is so engineered that it will transform one symbol into another if and only if the symbols stand in a certain semantic relation, say as premise to conclusion (under some interpretation), then we have the cognitivist resources to naturalistically (i.e., mechanistically) explicate the linkage between the causal properties of our neurophysiology and the semantic properties structuring our mental lives. “The idea that the brain is such a machine,” says Fodor “is the foundational hypothesis of Classical cognitive science” (Fodor and Pylyshyn 1988: 30).

In short, Fodor believes that the physical shape (syntax) of symbols making up our LOT links the causal properties of our neuroware (brain and CNS) to the semantic properties of our programs (minds) by virtue of an isomorphism of syntax and semantics presumably engineered by evolution. But, of course, it is obvious that Fodor is here working with a very anorexic notion of semantics. Semantics involves a whole lot more than mere truth-preservation, i.e., formal validity. What about semantic properties of reference and meaning? These semantic properties are notoriously intractable to formal reconstruction, and yet foundational to the natural languages in which we consciously configure our lives.

### **Fodor’s Ambivalence Towards Semantics**

Because Fodor relies so heavily on the serial symbol-manipulating operations of the classical digital computer to represent the kinematics of human intelligence, he often gives short shrift to the issue of *original* semantics. Computers running specific programs *already* have a rich semantics given to them by the programmer and user, and since Fodor believes he makes real progress in resolving some traditional problems in the philosophy of mind by relying on the computer for his cognitivist speculations, he feels no compunction for not having a naturalistic account of original semantics, how the internal mental representations of his LOT “have semantic properties in, one might say, the first instance. . . . a theory of how mental representations represent” (Fodor 1981: 31). As he openly admits, “one can do quite a lot of cognitive science without raising foundational - or, indeed, any - issues about the semanticity of mental representations” (Lycan [ed.], 1990: 315).

Fodor’s elusive comments on issues concerning the semanticity of mental representations (original semantics) are very interesting. He exclaims “Heaven only knows” what relation between him and things in the world makes it possible for him to think or refer to them (Fodor 1981: 253). In another place, he confesses to hoping for “someone very nice and clever” to turn up and show him how to provide his LOT with an original semantics (Fodor 1981: 223). At one point, he even considers that appealing to the propositional attitudes of personal agents might “be essential in explaining why the representation represents what it does,” but then he changes his mind, conceding that his

computational theory of mind might well “require us to view mental symbols as *sui generis*” (Lycan [ed.], 1990: 314). Despite his lack of a theory of original semantics, Fodor has some definite opinions on what an acceptable account of original semantics must look like. In fact, he is quite explicit on this score.

For one thing, an acceptable account *must* be a *naturalized* account: “a theory that articulates, in nonsemantic and nonintentional terms, sufficient conditions for one bit of the world to *be about* (to express, represent, or be true of) another bit” (Fodor 1988: 98). A naturalized account of original semantics, *if* it is going to be articulable in nonsemantic and nonintentional terms as Fodor requires (i.e., in terms that aren’t connected to the natural language through which the agent consciously thinks about his or her world), will have to be a *causal account*: i.e., an account that links, at a sub-personal level, proximal physical events in our neuroware to distal physical events in the world. According to Fodor, such an account would mean that “what makes it the case that (the Mentalese symbol) ‘water’ expresses the property H<sub>2</sub>O is that tokens of that symbol stand in certain *causal relations* to water samples” (Fodor 1988: 98: emphasis mine). Fodor recognizes that the naturalistic aspirations that led him to construe human intelligence as a species of computation require of him an accompanying construal of *original semantics as a species of causation*. But he seems blissfully unaware of the fact that even if the symbols of the LOT are *causally* linked to events in the world, this will not give them a semantics: effects do not *mean*, or *represent*, or *refer* to, their causes. But, if the relation of representation-to-represented cannot be naturalized through reconstruction in sub-personal terms, terms denoting only scientifically acceptable primitives, then his whole computational theory of human intelligence rests on a non-naturalized representational foundation.

That is, if the symbolic primitives of the LOT cannot be shown to represent certain properties in the world (instead of merely being effects of certain causes in the world) by virtue of some wholly natural process describable by science, then Fodor’s computational account of mind would be predicated on an almost dualist account of representation--symbols in the LOT would merely take on the role of Descartes’ postulated *res cogitans*: *je ne sais quoi* that somehow intrinsically represents. Clearly Fodor’s computational theory of mind does not sit easily with his representational account of the LOT, at least, not within the naturalistic framework in which he proposes them. The real problem that Fodor’s cognitivist program encounters and cannot get around is that meaning or representation requires slippage or a degree of indeterminacy between representation and represented where a personal agent can get in and bring the former (tacitly) to bear on the latter. The coupling of representation and represented must be *intentionally* forged, and this can only be brought about by the non-inferential, integrative acts of personal agents. But this is precisely what a computational construal of mental processes *cannot* accomplish, because it is pitched at a sub-personal level where the personal agent is analytically dissolved into inferential processes allegedly embodied in the causal transactions neurophysical events (see Polanyi 1958: 372).

## Summary

Fodor is committed to the view that inference is the essence of intelligence: that human intelligence is a knowledge-based, propositionally-structured, and inference-driven, virtual machine implemented by the physical-symbol system of human neurophysiology. On this view, intelligence’s essence is abstract, residing in symbolic and formal processes whose medium of implementation (hardware) is irrelevant; this is why Fodor can factor out all the somatic peculiarities of human mood, perception and motor skills and confidently assert “that there are information processing systems which share our psychology (instantiate its generalizations) but do not share our physical organization” (Fodor 1981: 9). That is, human intelligence has no *essential* bodily roots (cf. Polanyi 1966: 15). He

implicitly endorses a thesis of *disembodiment*, where intelligence bears no essential relation to any particular medium of traction in the world. Disembodied intelligence is purely formal intelligence, so its semantics only amount to preservation of theoretical consistency (validity). I believe this is why

Fodor's delineation of a slender semantics, is however, commensurate with the account of personal agents his commitments engender. Personal agents, within Fodor's cognitivist program, cannot but turn out to be epiphenomena supervening computational processes, i.e., ghosts in virtual machines: causally-feckless, and ontologically-suspect entities. But at least causally-feckless and ontologically-suspect entities could get by with the insubstantial semantics Fodor's computational program provides. We turn now to Polanyi's speculations on human intelligence.

## **Polanyi on the Emergence of Human Intelligence**

Polanyi's view of human intelligence derives from his recognition of its evolutionary past. He is convinced that we will only begin to understand the nature of human intelligence when we understand its historical antecedents—the human mind was not parachuted into a prestructured world. Intelligence, like everything else, evolved through encounters with an environment that afforded an ambiguous mixture of opportunities and risks. For Polanyi, that is, intelligence evolved as an embodied response to a challenging environment of significances, opportunities, affordances, meanings. Polanyi's metaphysical moorings are open to view in his fundamental premise that “meaning is the sort of thing the world is organized to bring about” (Polanyi and Prosch 1975: 182). Below, I shall, in brief and somewhat caricatured form, rehearse Polanyi's vision of mind's and meaning's progressive embodiment in the world.

In the final chapter of his (1958) *Personal Knowledge*, Polanyi shows us how meaning eventually emerged from physics. The world brought meaning out of physics by evolving purposive self-centred biological agents: “living individuals overcame the meaninglessness of the universe by establishing in it centres of subjective interests” (Polanyi 1958: 389). Through these “centres of subjective interest,” self-identical matter became duplicitous, taking a finite point of view on itself. And these centres of subjective interest eventually overcame their *mute* beasthood, creating “a new fabric of life *not* centred on individuals,” a form of life, that is, characterized by “universal intent,” whereby the universe gained an *articulate* point of view on itself, and meaning became embodied in an active centre (Polanyi 1958: 344). By creating and indwelling a symbolic means of intersubjective expression, self-centred bodily agents gradually, but stupendously, amplified the self-centred intelligence that evolution developed in their bodies, and thereby gave birth to the “noosphere,” transforming their mere biological agency into full-blown personhood (Polanyi 1958: 389). The human mind, according to Polanyi, “has been so far the ultimate stage in the awakening of the world” (Polanyi 1958: 405). The emergence of mind gave the universe's meaning an individualized articulate embodiment in molar agents who “formed societies [and] invented language” (Polanyi 1958: 388).<sup>3</sup>

Whether or not the following is exactly how Polanyi himself would have reconstructed the evolutionary twilight of man's emergence into articulate self-expression, I think it is in harmony with much of his speculations on anthropogenesis. The selective pressures on the vulnerable bodies of early hominids made broad and organized communal allegiances necessary to their survival. And these communal allegiances both fostered and depended upon certain expressive-mimetic gestures and sounds of intersubjective import. Over evolutionary time, the complexity of this pre-linguistic network of significant gestures increased: generation after generation added to this network new and more discriminating tokens of gestural significance until finally this dialectic of biology and proto-culture produced



a form of life whose social skills of survival loosened the biological grip of self-centred instincts enough to allow hominid evolution to transcend its purely biological identity through pressing itself symbolically into linguistic possibilities. This is what Polanyi refers to as “the second major rebellion against meaningless inanimate being” (the first major rebellion was, of course, “the rise of self-centred individuals”) (Polanyi 1958: 389). Man’s “rise from mute beasthood,” his transformation of bodily mimetic-gestural intelligence into articulated utterances of social significance, is “noogenesis”--the articulate embodiment of meaning in the human mind (Polanyi 1958: 389). When man participates in this life, his survival-oriented bodily intelligence

ceases to be merely an instrument of self-indulgence and becomes a condition of his calling. The inarticulate mental capacities developed in our body by the process of evolution become then the tacit coefficients of articulate thought. By the forming and assimilation of an articulate framework these tacit powers kindle a multitude of new intellectual passions. (Polanyi 1958: 389)

So, in summary, Polanyi conceives of the emergence of mind and meaning as a transformation of the purposive and expressive body coevally and dialectically enacted within a protocultural context by means of linguistic articulation. For Polanyi, the emergence of mind is a feat of evolution and can only be understood from an evolutionary perspective: mind emerged when hominid evolution produced an articulate social symbolism through which to express the meanings it created and discovered as it pressed itself into new and richer possibilities --possibilities of universal (not self-centred) intention, of eternal meaning (Polanyi 1958: 389).

But let’s turn from these speculations on anthropogenesis, and attend to Polanyi’s view of how a single individual child achieves a mental life, i.e., a mind. In other words, let’s shift from speculations on the phylogenesis of mind and meaning to a phenomenology of a mind’s ontogenesis within an *already in place* cultural context of socially-sustained linguistic meanings. This will help us see how, on Polanyi’s view, the mind and original semantics are intimately connected to a child’s acquisition of a first language: a mother tongue.

## **Polanyi on the Development of a Mind**

According to Polanyi, we, the benefactors of millions of years of evolution, come to embody a mind when we learn to indwell our language community’s socially-sustained network of meanings: i.e., its articulate framework:

We come into existence mentally by adding an articulate framework to our bodies. Human thought grows only within language and since language can exist only in society, all thought is rooted in society (Hall 1968: 67).

The child grafts an articulate system of expression onto “the inarticulate mental capacities developed in [its] body by the process of evolution”; the child’s body provides then, “the tacit coefficients of [its] articulate thought” (Polanyi 1958: 389). A mind emerges, therefore, when a young child’s mute bodily passions (passions shared with its mammalian ancestors) are gradually channelled through, and transfigured by, the network of social commitments and normative constraints embodied in, and mediated through, its mother tongue: a purposive body is transformed into an intentional agent.<sup>4</sup> Through learning to indwell the articulate framework of its mother tongue, a child’s sub-linguistic, pre-propositional, and a-critical bodily intelligence is transformed into a symbolically-articulated, and propositionally-structured, rational mind. The emergence of a child’s mind is, therefore, a biosocial achievement: the child’s natural

bodily passions (appetitive, motoric, and perceptive drives) acquire the skill of linguistic expression (and become thereby mental passions) through the training and shaping imposed on them by a linguistic community.

On Polanyi's view, the individual child comes into existence mentally *always already* enconced in a nurturing linguistic community whose idiom functions as the child's *consciously* deployed "language of thought" (Schwartz [ed.] 1974: 75). That is, propositional thought begins as conscious and intentional public conversation, not as unconscious and private computations. Polanyi would sharply disagree with Fodor's view that thought is propositionally-structured all the way down. Polanyi's evolutionary perspective has propositionally-structured intelligence resting on a pre-propositional foundation of indwelling: The emergence of the mind is, therefore, predicated on a child's assimilation of a pre-existent articulate framework (i.e., a natural language), which, in turn, is predicated on the child's *natural* bodily capacity to acquire new skills.

"To use language," Polanyi submits, "is to extend our bodily equipment and become intelligent human beings" (Grene [ed.] 1969: 148). Polanyi recognizes that natural bodily drives and passions "are refashioned and amplified into something new by words" (Polanyi 1958: 194). By pouring our mute bodily passions into the articulate framework of our mother tongue, we develop new discriminative faculties and sensitivities, because "*as each of us interiorizes our cultural heritage*, he grows into a person seeing the world and experiencing life in terms of this outlook" (Grene [ed.] 1969: 148; emphasis mine). Our first language embodies the network of presuppositions, categories, and concepts under the constraints of which our mind emerges (Polanyi 1958: 266-67). As a child gradually comes to indwell a mother tongue, his innate *répertoire* of bodily expression acquires a cultural dimension that is grounded in its mother tongue's semantics: he participates in the semantic content of his cultural heritage, thus extending his bodily expressivity into the noosphere where he can perform speech-acts and become a responsible participant in his society.

Once a child has learned to indwell her mother tongue, she no longer responds to stimuli *simpliciter*, but encounters objects *under descriptions*. That is, her point of view becomes impregnated by the intensional semantics of her mother tongue. Her sense-bound memory is supplemented and penetrated by a verbal memory, and she soon gains the skill of making sense of her experiences by re-describing them and weaving them into coherent narratives. Moreover, she gains access to an explicit domain of possibility and counterfactuality: she can now refer to her past, fear her future, care about her present, tell lies, make her own desires and beliefs objects of her own (second-order) desires and beliefs, and even dread her own non-existence. A whole multitude of intellectual passions are kindled in a child when she assimilates the articulate framework of her culture, enabling her to participate in the rights and responsibilities of the noosphere (Polanyi 1959: 60).

So, for Polanyi, mind is neither *res cogitans* nor software driven by some hardware's syntax, but an articulate way of being *in* the world.<sup>5</sup> And if this is the case, then Fodor's Menoesque problem of first language acquisition (i.e., we must first know a language in order to learn one) is *really* reduced to the problem of how we acquire the ability to indwell external entities. Polanyi's account of indwelling obviates the need for the innate LOT computations. Indwelling is a *basic* bodily action that Polanyi traces to "the inarticulate mental capacities developed in our body by the process of evolution" (Polanyi 1958: 389). Indwelling *isn't*, therefore, something we learn, or acquire, but our very means of learning or acquiring any skill whatsoever. In contrast to Fodor's postulated computations in the LOT (sub-personal activities that our CNS does), indwelling is something we do, but *not* by virtue of doing anything else. But this does not mean, of course, that we cannot indwell some thing *intentionally*, but only that *there is always a measure of irreducible indwelling preceding every intention*. Indwelling is primitive and phenomenologically simple: it is not

something we can bring about by doing something else first, because it is the *conditio sine qua non* of doing anything. However, we can *learn* to indwell things like hammers, skillful performances, physiognomies, or *first languages*, but the indwelling itself isn't something we can learn or teach--we can only *do* it. In the place of Fodor's hard-wired, universal, and propositionally-structured LOT in which a child purportedly forms hypotheses about the extensions of natural language predicates and then tests them, Polanyi has our first-language acquisition bottom-out in the basic actions of our universal, but sub-linguistic and pre-propositional bodily intelligence:

The child's way of learning to speak from his adult guardians is . . . akin to the young mamal's and young bird's mimetic responses to its nurturing, protecting, and guiding seniors (Polanyi 1958: 206).

“Indwelling is,” Polanyi claims, “being-in-the-world” (Polanyi 1964: xi). It is the way we are embodied as points of view in, and points of action on, the world; by its means “we keep expanding our body into the world” forming thereby “an interpreted universe” (Polanyi 1966: 29). Indwelling modifies our own limitations and possibilities within the world. And when we, as prelinguistic children, first indwell a mother tongue our identity undergoes a radical change because we are irreversibly taken out of a mere environment of instinct and stimuli and plugged into a whole network of socially-sustain symbols, meanings and practices accomplished by those who, before us, have taken the step into an articulate way of being-in-the-world (Polanyi 1958: 266-67):

Every time we assimilate a tool to our body our identity undergoes some change; our person expands into new modes of being. I have shown before that the whole realm of human intelligence is grounded on the use of language. . . . all mental life by which we surpass the animals is evoked in us as we assimilate the articulate framework of our culture (Polanyi 1959: 31).

From this perspective, first-language acquisition is predicated upon a child's apprenticeship in the practices and institutions of a language community. And, of course, apprenticeship in a language community can come about only because a child is always already “equipped by nature” to *indwell* the physiognomy and behaviors of its caretakers (Grene [ed.] 1969: 220). So while Fodor stems an infinite regress of languages by conceiving first-language acquisition as a matter of a child's unconscious but highly abstract operations on explicit symbols within its LOT, Polanyi's account of first-language acquisition stems the infinite regress by bottoming-out in the tacit dimension: in a primitive (basic), non-linguistic and innate bodily action, *indwelling*.

Our task in the next section is to see how Polanyi's conception of indwelling as the *conditio sine qua non* of mind's embodiment and meaning's emergence ramifies into a critique of Fodor's CS.

## Polanyi on Semantics

According to Polanyi, meaning is embodied in acts of indwelling whereby a personal agent extends itself into the world by integrating subsidiary items into a sustainable focus of attention. And *linguistic* meaning obtains by virtue of human agents indwelling *linguistic* tokens as a means of self- and intersubjective-expression. Polanyi will

have nothing to do with notions of semantics that aren't directly tied to the activities of human agency (Polanyi 1959: 22): e.g., Fodor's idea that representation can somehow be reduced to the sub-personal, to event-causation. On Polanyi's view, the semantic properties of our thought derive from our informal acts of indwelling whereby we bring the linguistic tokens (representations) of our mother tongue to bear tacitly on our focal targets (the represented). Consequently, Polanyi has no need to postulate a LOT whose primitive neurophysically-embodied symbols function as semantic primitives.

In his (1967) essay, "Sense-giving and Sense-reading," Polanyi speaks of language's "informal semantic structure" (Greene [ed.]1969: 181). He unpacks this structure as that of *indwelling*: where a "meaningful relation of a subsidiary to a focal is formed by the action of a person who integrates one to the other, and the relation persists by the fact that the person keeps up this integration" (Greene [ed.]1969: 182; emphasis mine). In a different essay, "Logic and Psychology," Polanyi is careful to point out that subsidiary and focal items are not "linked together of [their] own accord" (Polanyi 1968: 30). That is, meaning is not causally, but *intentionally* instantiated. So linguistic meaning comes into being and persists only by virtue of persons attending *from* linguistic tokens *to* their focal import. The personal agent is ineliminable from Polanyi's construal of semantics: "nothing ... can ever mean anything in itself." All semantic functions, he submits, "are the tacit operations of a person" (Polanyi 1959: 22).

In 1968, Polanyi was interviewed for the first issue of *Psychology Today*. Mary Harrington Hall asked him to comment on contemporary linguistics. He responded, noting that Chomsky's strictly formalist approach to linguistics (an approach that Fodor has taken-up and extended) leaves one with no real means to talk about *meaning*. The only way to cope with linguistic meaning, Polanyi claims, is to recognize that *meaning is a relational reality*, that meaning is "the relationship between the subsidiary and that on which it bears" (Hall 1968: 67). And applying this point to the issue of computer intelligence, Polanyi notes:

This is why *meaning cannot be introduced by a computer*, because the computer can only operate with focally known elements. It can never reproduce two different levels of awareness (Hall 1968: 67; emphasis mine).

That is, there is no personal agent whose integrative acts generate subsidiary awareness. Only an embodied agent who has access to the world by virtue of attending *from* its body bifurcates its awareness into two levels: the subsidiary and the focal. A computer merely opens and closes logical gates (micro-switches) according to constraints the programmer has imposed on them; it doesn't attend *from* at all. So its "world" is merely notional, a series of focal data-structures programmed into its electronically-driven binary code. But if this is so, how does one explain the semantics of the computer? How does one account for the fact a computer's computations really are sensitive to some semantic distinctions?

Notice in the quotation above that Polanyi didn't say that meaning cannot be ascribed to a computer, but only that "meaning cannot be *introduced* by a computer." That is, *original semantics cannot be generated by computers*, but only borrowed. Polanyi argues that "formal systems of symbols [representations] and operations [rules]" depend for their meaning on the "unformalized operations" that are "performed by a person with the aid of the formal system, when the person relies on its use" (Polanyi 1958: 258). That is, whether a formalized system of representation is

automated or not, its semantics derive from the person(s) using it. Computers then, on this construal, exhibit what Polanyi calls a “necessary relatedness to persons” which is a property that essentially restricts the independence of a machine, and reduces the status of automata below that of thinking persons:

For a machine is a machine only for someone who relies on it (actually or hypothetically) for some purpose, that he believes to be attainable by what he considers to be the proper functioning of the machine: it is the instrument of a person who relies on it. This is the difference between machine and mind. A man’s mind can carry out feats of intelligence by aid of a machine and also *without* such aid, while a machine can function only *as the extension of a person’s body* . . . (Polanyi 1958: 262; emphasis mine).

The human programmer is the *essential* link between the computer’s internal states and the objects or concepts they are deployed to denote. And it is only because human computer users *take* these operations to denote what the programmer designed them to denote that the computer warrants the ascription of computing, rather than say, heating the room, or providing a pleasant background hum. Computers can function as computers only “as extensions of a person’s body” because all meaning is *introduced* by persons treating external things as extensions of their bodies. On Polanyi’s view, human bodies are the matrices of original semantics; they are the from-which of every intention (Grene [ed.] 1969: 183-184).

So the computer’s semantics must be a *borrowed semantics*, a semantics bestowed on its symbolism by its human programmers and users: “the symbol can be conceived as such only in the eyes of the person who relies on [it] to achieve or to signify something” (Polanyi 1958: 61). But, Polanyi’s account of indwelling does, however, demonstrate why the computer may appear to be a syntax-driven *semantic* engine. Since indwelling tacitly imbues the indwelt with a bearing on a focus, the computer’s operations appear to be possessed of an intrinsic meaning: by using them to extend our focal awareness, we indwell them and automatically and tacitly supplement their symbols and rules with a bearing on our focal concern. And “this act goes unnoticed” (Grene [ed.] 1969: 151; c.f., Polanyi 1958: 169), because indwelling’s self-effacing transitivity renders the agent’s activities phenomenologically transparent, transposing them into the meaning of the agent’s focal object (Polanyi 1958: 60-61); this is the “semantic aspect” of tacit knowing (Polanyi 1966: 12-13), and the dynamic of self-forgetfulness behind Fodor’s inveterate tendency to commit the fallacy of misplaced semantics: assigning independent semantic properties to his postulated sub-personal atoms of cognition.

From this Polanyian perspective, Fodor’s aim of accounting for the semantics of human cognition by appealing to the computer model is preposterous: whatever semantic properties a computer may be said to possess derive, in the first place, from the informal acts of tacit knowing underlying all human cognition. This is why Polanyi calls the attempt to reduce human thought to formal processes “self-contradictory.” “The pursuit of formalization,” says Polanyi, “will find its true place in a tacit framework” (Grene [ed.] 1969: 156).

## Conclusion

Fodor's CS weds a Kantian epistemology to an empiricist philosophy of mind. Kant's *a priori* categories have become Fodor's rules, and the empiricist's impressions have become his primitive representations. But, whereas Kant was forced to postulate a transcendental subject in whose cogitations the rules were embodied, and the empiricists were forced to recognize their implicit reliance on an internal homunculus who viewed and associated the impressions, Fodor, in his desire to naturalize mind, seeks to obviate the need for a central agent by decomposing into dumber and dumber activities all the intelligent activities ascribed to it, until a point is reached that a series of micro-switches could perform its tasks. But Polanyi's account of mind's emergence and meaning's tacit dynamics reveals that, in reality, Fodor can only entertain the prospects of a naturalization of human intelligence because he is blissfully unaware of his own tacit contribution of semantics to the computational operations he appeals to in his herculean effort to explain human intelligence in sub-personal terms.

It is not just a little bit ironic that, in Fodor's attempt to naturalize human cognition by extending the mechanistic metaphor - in his attempt, that is, to not be anthropomorphic about man himself--he ends up creating a *new* breed of anthropomorphism: projecting the mechanistic (algorithmic) properties of a *human artifact* (virtual machine) onto nature (neurophysiology) and then reading these properties into the human psyche as if they explained the intelligence that invented this artifact in the first place. From a Polanyian point of view, Fodor's naturalization project is but another example of the "crippling mutilations" (Polanyi 1958: 381) that objectivism forces on our picture of human intelligence: it requires nothing less than "a specifiably functioning mindless knower" (Polanyi 1958: 264). Surely Fodor's work confirms Polanyi's claim that: "Any attempt rigorously to eliminate our human perspective from our picture of the world must lead to absurdity" (Polanyi 1958: 3).

## ENDNOTES

1. It is important to realize that the LOT is not the language in which we think; it is the language in which our *neurophysiology* computes, but Fodor takes a subset of these computations to be token-identical to human thinking, see Fodor 1975: 49ff.

2. It is interesting to note that these unconscious and subpersonal doings of the child's neurophysically-embodied LOT is what Fodor refers to as tacit knowledge. Tacit knowledge, for Fodor, is the program that underlies an organism's behavior (Fodor 1981: 78).

3. For some interesting speculations on how the dialectic of hominid biological vulnerability and hazardous environment occasioned the need for social organization, and how this, in turn, fostered and sustained the emergence of an articulate framework, see John McCrone (1990) and Merlin Donald (1991). Note, however, that both, at times, are rather naive in their deployment of the computer metaphor.

4. For more details on how this transformation might come about, see my forthcoming "The Body Comes All the Way Up," *International Philosophical Quarterly*, (Spring 1994).

5. That is, humans of mental competence are not in the world like an object enclosed in some physical space and bearing only external relations (causal and spatial) to other objects. Rather, humans are in the world in the metaphorical sense of “being in the midst of completing the first draft of a paper.” Humans of mental competence are bodily points of view on and points of action in the world: agents of concerned relations who are centrally situated in their phenomenal field and whose experience of the world is penetrated by the categories and concepts of their mother tongue. (See Polanyi’s Heideggerian gloss on indwelling, Polanyi 1964: x-xi)

## WORKS CITED

- Donald, Merlin. (1991) *Origins of the Modern Mind*, Cambridge, Mass.: Harvard University Press.
- Fodor, Jerome. (1975) *The Language of Thought*, Cambridge, Mass.: MIT Press.
- . (1981) *Representations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge, Mass.: MIT Press.
- . (1985) "Fodor's Guide to Mental Representation," *Mind* 94; 76-100.
- . (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, Mass.: MIT Press.
- Fodor, Jerome & Zeno Pylyshyn. (1988) "Connectionism and Cognitive Architecture: A Critical Analysis," *Cognition* 28,3-71.
- Grene, Marjorie (ed.). (1969) *Knowing and Being: Essays by Michael Polanyi*, Chicago: University of Chicago Press.
- Hall, Mary Harrington. (1968) "A Conversation with Michael Polanyi," *Psychology Today* 1;20-25 & 65-67.
- Haugeland, John. (1985) *Artificial Intelligence: The Very Idea*, Cambridge, Mass.: MIT Press.
- Lycan, William G. (ed.). (1990) *Mind and Cognition*, Cambridge, Mass.: Basil Blackwell, Inc.
- McCrone, John. (1991) *The Ape that Spoke*, New York: William Morrow and Company, Inc.
- Polanyi, Michael. (1958) *Personal Knowledge: Towards a Post-Critical Philosophy*, Chicago: University of Chicago Press.
- . (1959) *The Study of Man*, Chicago: University of Chicago Press.
- . (1964) "Preface to the Torchbook Edition," *Personal Knowledge*, New York: Harper & Row; ix-xi.
- . (1966) *Tacit Dimension*, Garden City: Doubleday & Company.
- . (1968) "Logic and Psychology," *American Psychologist* 23; 27-43.
- . (1969) "On Mind and Body," *The New Scholasticism* 43; 195-204.
- Polanyi, Michael & Harry Prosch. (1975) *Meaning*, Chicago: University of Chicago Press.
- Pylyshyn, Zenon W. (1984) *Computation and Cognition*, Cambridge, Mass: MIT Press.
- Schwartz, Fred (ed.) (1974) *Scientific Thought and Social Reality: Essays by Michael Polanyi*, New York: International University Press.