Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian

Hubert L. Dreyfus

I. The Convergence of Computers and Philosophy

When I was teaching at MIT in the early sixties, students from the Artificial Intelligence Laboratory would come to my Heidegger course and say in effect: "You philosophers have been reflecting in your armchairs for over 2000 years and you still don't understand intelligence. We in the AI Lab. have taken over and are succeeding where you philosophers have failed. We are now programming computers to exhibit human intelligence: to solve problems, to understand natural language, to perceive, and to learn."[1] In 1968 Marvin Minsky, head of the AI lab, proclaimed: "Within a generation we will have intelligent computers like HAL in the film, 2001."[2]

As luck would have it, in 1963, I was invited by the RAND Corporation to evaluate the pioneering work of Alan Newell and Herbert Simon in a new field called Cognitive Simulation CS). Newell and Simon claimed that both digital computers and the human mind could be understood as physical symbol systems, using strings of bits or streams of neuron pulses as symbols representing the external world. Intelligence, they claimed, merely required making the appropriate inferences from these internal representations. As they put it: "A physical symbol system has the necessary and sufficient means for general intelligent action."[3]

As I studied the RAND papers and memos, I found to my surprise that, far from replacing philosophy, the pioneers in CS and AI had learned a lot, directly and indirectly from the philosophers. They had taken over Hobbes' claim that reasoning was calculating, Descartes' mental representations, Leibniz's idea of a "universal characteristic" – a set of primitives in which all knowledge could be expressed, -- Kant's claim that concepts were rules, Frege's formalization of such rules, and Wittgenstein's postulation of logical atoms in his Tractatus. In short, without realizing it, AI researchers were hard at work turning rationalist philosophy into a research program.

At the same time, I began to suspect that the insights formulated in existentialist armchairs, especially Heidegger's and Merleau-Ponty's, were bad news for those working in AI laboratories-- that, by combining representationalism, conceptualism, formalism, and logical atomism into a research program, AI researchers had condemned their enterprise to confirm a failure.

II. Symbolic AI as a Degenerating Research Program

Using Heidegger as a guide, I began to look for signs that the whole AI research program was degenerating. I was particularly struck by the fact that, among other troubles, researchers were running up against the problem of representing significance and relevance – a problem that Heidegger saw was implicit in Descartes' understanding of the world as a set of meaningless facts to which the mind assigned what Descartes called values and John Searle now calls function predicates.

But, Heidegger warned, values are just more meaningless facts. To say a hammer has the function of being for hammering leaves out the defining relation of hammers to nails and other equipment, to the point of building things, and to our skills. – all of which Heidegger called readiness-to-hand --and so attributing functions to brute facts couldn't capture the meaningful organization of the everyday world. "[B]y taking refuge in 'value'-characteristics," Heidegger said, "we are … far from even catching a glimpse of being as readiness-to-hand ..."[4]

Minsky, unaware of Heidegger's critique, was convinced that representing a few million facts about objects including their functions, would solve what had come to be called the commonsense knowledge problem. It seemed to me, however, that the real problem wasn't storing millions of facts; it was knowing which facts were relevant in any given situation. One version of this relevance problem is called the frame problem. If the computer is running a representation of the current state of the world and something in the world changes, how does the program determine which of its represented facts can be assumed to have stayed the same, and which might have to be updated?

As Michael Wheeler in his recent book, Reconstructing the Cognitive World, puts it:

> [G]iven a dynamically changing world, how is a nonmagical system ... to take account of those state changes in that world ... that matter, and those unchanged states in that world that matter, while ignoring those that do

> not? And how is that system to retrieve and (if necessary) to revise, out of all the beliefs that it possesses, just those beliefs that are relevant in some particular context of action?[5]

Minsky suggested that, to avoid the frame problem, AI programmers could use descriptions of typical situations like going to a birthday party to list and organize those, and only those, facts that were normally relevant. Perhaps influenced by a computer science student who had taken my phenomenology course, Minsky suggested a structure of essential features and default assignments-- a structure Husserl had already proposed and called a frame.[6]

But a system of frames isn't *in* a situation, so in order to select the possibly relevant facts in the current situation one would need frames for recognizing situations like birthday parties, and for telling them from other situations such as ordering in a restaurant. But how, I wondered, could the computer select from the supposed millions of frames in its memory the relevant frame for selecting the birthday party frame as the relevant frame, so as to see the current relevance of an exchange of gifts? It seemed to me obvious that any AI program using frames to organize millions of meaningless facts so as to retrieve the currently relevant ones was going to be caught in a regress of frames for recognizing relevant frames for recognizing relevant facts, and that, therefore, the commonsense knowledge storage and retrieval problem wasn't just a problem but was a sign that something was seriously wrong with the whole approach.

Unfortunately, what has always distinguished AI research from a science is its failure to face up to and learn from its failures. In this case, to avoid facing the relevance problem the AI programmers at MIT in the sixties and early seventies limited their programs to what they called micro-worlds – artificial situations in which the small number of features that were possibly relevant was determined beforehand. Since this approach obviously avoided the real-world frame problem, PhD students were compelled to claim in their theses that their micro-worlds could be made more realistic, and that the techniques they introduced could be generalized to cover commonsense knowledge. There were, however, no successful follow-ups.[7]

The work of Terry Winograd is typical. His "blocks-world" program, SHRDLU, responded to commands in ordinary English instructing a virtual robot arm to move blocks displayed on a computer screen. It was the parade case of a micro-world program that really worked – but of course only in its micro-world. So to develop the expected generalization of his techniques, Winograd started working on a new Knowledge Representation Language. (KRL). His group, he said, was "concerned with developing a formalism, or 'representation,' with which to describe ... knowledge." And he added: "We seek the 'atoms' and 'particles' of which it is built, and the 'forces' that act on it."[8]

But his approach wasn't working either. Indeed, Minsky has recently acknowledged in *Wired Magazine* that AI has been brain dead since the early 70s when it encountered the problem of commonsense knowledge. [9] Winograd, however, unlike his colleagues, was courageous enough to try to figure out what had gone wrong. So in the mid 70ies he began having weekly lunches with John Searle and me to discuss his problems in a broader philosophy context. Looking back, Winograd says: "My own work in computer science is greatly influenced by conversations with Dreyfus over a period of many years."[10]

After a year of such conversations, and after reading the relevant texts of the existential phenomenologists, Winograd abandoned work on KRL and began including Heidegger in his Computer Science courses at Stanford. In so doing, he became the first high-profile deserter from what was, indeed, becoming a degenerating research program. John Haugeland now refers to the symbolic AI of that period as Good Old Fashioned AI —GOFAI for short—and that name has been widely accepted as capturing its current status. Michael Wheeler argues explicitly that a new paradigm is already taking shape. He maintains:

> [A] Heideggerian cognitive science is … emerging right now, in the laboratories and offices around the world where embodied-embedded thinking is under active investigation and development.[11]

Wheeler's well informed book could not have been more timely since there are now at least three versions of supposedly Heideggerian AI that might be thought of as articulating a new paradigm for the field: Rodney Brooks' behaviorist approach at MIT, Phil Agre's pragmatist model, and Walter Freeman dynamic neural model. All three approaches accept Heidegger's critique of Cartesian internalist representationalism, and, instead, embrace John Haugeland's slogan that cognition is embedded and embodied.[12]

III. Heideggerian AI, Stage One: Eliminating representations by building Behavior-based Robots

Winograd sums up what happened at MIT after he left for Stanford.

> For those who have followed the history of artificial intelligence, it is ironic that [the MIT] laboratory should become a cradle of "Heideggerian AI." It was at MIT that Dreyfus first formulated his critique, and, for twenty years, the intellectual atmosphere in the AI Lab was overtly hostile to recognizing the implications of what he said. Nevertheless, some of the work now being done at that laboratory seems to have been affected by Heidegger and Dreyfus. [13]

Here's how it happened. In March l986, the MIT AI Lab under its new director, Patrick Winston, reversed Minsky's attitude toward me and allowed, if not encouraged, several graduate students, led by Phil Agre and John Batali, to invite me to give a talk.[14] I called the talk, "Why AI Researchers should study Being and Time." In my talk I repeated what I had written in l972 in What Computers Can't Do: "[T] he meaningful objects ... among which we live are not a *model* of the world stored in our mind or brain; *they are the world itself."[15]* And I quoted approvingly a Stanford Research Institute report that pointed out that, "It turned out to be very difficult to reproduce in an internal representation for a computer the necessary richness of environment that would give rise to interesting behavior by a highly adaptive robot,"[16] and concluded that "this problem is avoided by human beings because their model of the world is the world itself."[17]

The year of my talk, Rodney Brooks, who had moved from Stanford to MIT, published a paper criticizing the GOFAI robots that used representations of the world and problem solving techniques to plan their movements. He reported that, based on the idea that "the best model of the world is the world itself," he had "developed a different approach in which a mobile robot uses the world itself as is own representation –

continually referring to its sensors rather than to an internal world model."[18] Looking back at the frame problem, he says:

> And why could my simulated robot handle it? Because it was using the world as its own model. It never referred to an internal description of the world that would quickly get out of date if anything in the real world moved.[19]

Although he doesn't acknowledge the influence of Heidegger directly,[20] Brooks gives me credit for "being right about many issues such as the way in which people operate in the world is intimately coupled to the existence of their body."[21]

Brooks' approach is an important advance, but Brooks' robots respond only to *fixed features* of the environment, not to context or changing significance. They are like ants, and Brooks aptly calls them "animats." Brooks thinks he does not need to worry about learning. He proposes it as a subject for future research but not currently his concern.[22] But by operating in a fixed world and responding only to the small set of possibly relevant features that their receptors can pick up, Brooks' animats beg the question of changing relevance and so finesse rather than solve the frame problem.

Merleau-Ponty's work, on the contrary, offers a nonrepresentational account of the way the body and the world are coupled that suggests a way of avoiding the frame problem. According to Merleau-Ponty, as an agent acquires skills, those skills are "stored", not as representations in the mind, but as a bodily readiness to respond to the solicitations of situations in the world. What the learner acquires through experience is not *represented* at all but is *presented* to the learner as more and more finely discriminated situations, and, if the situation does not clearly solicit a single response or if the response does not produce a satisfactory result, the learner is led to further refine his discriminations, which, in turn, solicit more refined responses. For example, what we have learned from our experience of finding our way around in a city is sedimented in how that city *looks* to us. Merleau-Ponty calls this feedback loop between the embodied agent and the perceptual world the *intentional arc*. He says: "Cognitive life, the life of desire or perceptual life – is subtended by an 'intentional arc' which projects round about us our past, our future, [and] our human setting."[23]

Brooks comes close to a basic existentialist insight spelled out by Merleau-Ponty,[24] viz. that intelligence is founded on and presupposes the more basic way of coping we share with animals, when he says:

> The "simple" things concerning perception and mobility in a dynamic environment … are a necessary basis for "higher-level" intellect. …Therefore, I proposed looking at simpler animals as a bottom-up model for building intelligence. It is soon apparent, when "reasoning" is stripped away as the prime component of a robot's intellect, that the dynamics of the interaction of the robot and its environment are primary determinants of the structure of its intelligence.[25]

Brooks is realistic in describing his ambitions and his successes. He says:

> The work can best be described as attempts to emulate insect-level locomotion and navigation. …There have been some behavior-based attempts at exploring social interactions, but these too have been modeled after the sorts of social interactions we see in insects.[26]

Surprisingly, this modesty did not deter Brooks and Daniel Dennett from repeating the extravagant optimism characteristic of AI researchers in the sixties. As in the days of GOFAI, on the basis of Brooks' success with insect-like devices, instead of trying to make, say, an artificial spider, Brooks and Dennett decided to leap ahead and build a humanoid robot. As Dennett explained in a l994 report to The Royal Society of London:

> A team at MIT of which I am a part is now embarking on a long-term project to design and build a humanoid robot, Cog, whose cognitive talents will include speech, eye-coordinated manipulation of objects, and a host of self-protective, self-regulatory and self-exploring activities.[27]

Dennett seems to reduce this project to a joke when he adds in all seriousness: "While we are at it, we might as well try to make Cog crave human praise and company and even exhibit a sense of humor."[28] (That should have been my put down line.)

Of course, the "long term project" was short lived. Cog failed to achieve any of its goals and is already in a museum.[29] But, as far as I know, neither Dennett nor anyone connected with the project has published an account of the failure and asked what mistaken assumptions underlay their absurd optimism. In response to my asking what had been learned, Dennett offered one of the usual AI lame excuses for failure-- in this case, the lack of graduate students—and put the usual misleading positive spin on what had been accomplished:

> Cog never advanced beyond the toddler stage in any competence (and never got out of neonate in many others). But then, after the first few years, only two or three grad students were working on it full time. *Progress was being made on all the goals, but slower than had been anticipated.*[30]

If progress was actually being made the graduate students wouldn't have left, or others would have continued to work on the project. Clearly some specific assumptions must have been mistaken, but all we find in Dennett's assessment is the implicit assumption that human intelligence is on a continuum with insect intelligence, and that therefore adding a bit of complexity to what has already been done with animats counts as progress toward humanoid intelligence. At the beginning of AI research, Yehoshua Bar-Hillel called this way of thinking the first-step fallacy, and my brother quipped, "it's like claiming that the first monkey that climbed a tree was making progress towards flight to the moon."

Compared to Dennett's conviction that Brooks' AI research is progressing along a continuum that will eventually lead from animats to humanly intelligent machines, Brooks is prepared to entertain the possibility that he is barking up the wrong tree. He concludes a discussion of his animats with the sober comment that:

> Perhaps there is a way of looking at biological systems that will illuminate an inherent necessity in some aspect of the interactions of their parts that is completely missing from our artificial systems…. I am not suggesting that we need go outside the current realms of mathematics, physics, chemistry, or biochemistry. Rather I am suggesting that perhaps at this point we simply do not *get it*, and that there is some fundamental change necessary in our thinking in order that we might build artificial systems

> that have the levels of intelligence, emotional interactions, long term stability and autonomy, and general robustness that we might expect of biological systems.[31]

We can already see that Heidegger and Merleau-Ponty would say that, in spite of the breakthrough of giving up internal symbolic representations, Brooks, indeed, doesn't get it -- that what AI researchers have to face and understand is not only why our everyday coping couldn't be understood in terms of inferences from symbolic representations, as Minsky's intellectualist approach assumed, but also why it can't be understood in terms of responses caused by fixed features of the environment, as in Brooks' empiricist approach. AI researchers need to consider the possibility that embodied beings like us take as input energy from the physical universe and respond in such a way as to open them to a world organized in terms of their needs, interests, and bodily capacities, without their *minds* needing to impose meaning on a meaningless given, as Minsky's frames require, nor their *brains* converting stimulus input into reflex responses, as in Brooks' animats.

At the end of this talk, I'll suggest that Walter Freeman's neurodynamics offers a radically new Merelau-Pontian approach to human intelligence – an approach compatible with physics and grounded in the neuroscience of perception and action. But first we need to examine another approach to AI contemporaneous with Brooks' that actually calls itself Heideggerian.

IV. Heideggerian AI, Stage 2: Programming the Ready-to-hand

In my talk at the MIT AI Lab, I not only introduced Heidegger's non-representational account of the relation of Dasein (human being) and the world, I also explained that Heidegger distinguished two modes of being: the *readiness-to-hand* of equipment when we are involved in using it, and the *presence-at-hand* of objects when we contemplate them. Out of that explanation and the lively discussion that followed, grew the second type of Heideggerian AI. The first to acknowledge its lineage.

This new approach took the form of Phil Agre's and David Chapman's program, *Pengi*, which guided a virtual agent playing a computer game called Pengo, in which the player and penguins kick large and deadly blocks of ice at each other.[32] Agre's approach, which he called "interactionism," was more self-consciously Heideggerian than Brooks, in that Agre proposed to capture what he calls "Heidegger's account of everyday routine activities." [33]

In his book, Computation and Human Experience, Agre takes up where my talk left off, saying:

> I believe that people are intimately involved in the world around them and that the epistemological isolation that Descartes took for granted is untenable. This position has been argued at great length by philosophers such as Heidegger and Merleau-Ponty; I wish to argue it technologically. [34]

Agre's interesting new idea is that the world of the game in which the Pengi agent acts is made up, not of present-at-hand facts and features, but of possibilities for action that require appropriate responses from the agent. To program this involved approach Agre used what he called "deictic representations." He tells us:

> This proposal is based on a rough analogy with Heidegger's analysis of everyday intentionality in Division I of *Being and Time*, with objective intentionality corresponding to the present-at-hand and deictic intentionality corresponding to the ready-to-hand.[35]

And he explains:

> [Deictic representations] designate, not a particular object in the world, but rather a role that an object might play in a certain time-extended pattern of interaction between an agent and its environment. Different objects might occupy this role at different times, but the agent will treat all of them in the same way.[36]

Looking back on my talk at MIT and rereading Agre's book, I now see that, in a way, Agre understood Heidegger's account of readiness-to-hand better than I did at the time. I thought of the ready-to-hand as a special class of *entities*, viz. equipment, whereas the Pengi program treats what the agent responds to purely as *functions*. For Heidegger and Agre the ready-to-hand is not a *what* but a *for-what*. [37]

As Agre saw, Heidegger wants to get at something more basic than simply a class of objects defined by their use. At his best, Heidegger would, I think, deny that a hammer in a drawer has readiness-to-hand as its way of being. Rather, he sees that, *for the user*, equipment is encountered as a solicitation to act, not an entity with function features. He notes that: "When one is wholly devoted to something and 'really' busies oneself with it, one does not do so just alongside the work itself, or alongside the tool, or alongside both of them 'together'."[38] And he adds: "the peculiarity of what is proximally ready-to-hand is that, in its readiness-to-hand, it must, as it were, withdraw in order to be ready-to-hand quite authentically."[39]

As usual with Heidegger, we must ask: what is the phenomenon he is pointing out? In this case he sees that, to observe our hammer or to observe ourselves hammering undermines our skillful coping. We can and do observe our surroundings while we cope, and sometimes, if we are learning, monitoring our performance as we learn improves our performance in the long run, but in the short run such attention interferes with our performance. For example, while biking we can observe passers by, or think about philosophy, but if we start observing how we skillfully stay balanced, we risk falling over.

Heidegger struggles to describe the special, and he claims, basic, way of being he calls the ready-to-hand. The Gestaltists would later talk of "*solicitations*", and J.J. Gibson, even later, would introduce the idea of "*affordances*." In Phenomenology of Perception Merleau-Ponty speaks of "*motivations*" and later, of "*the flesh*." All these terms point at what is not objectifyable--a situation's way of drawing one into it.

In his 1925 course, Logic: The Question of Truth Heidegger describes our most basic experience of what he later calls "pressing into possibilities" not as dealing with the desk, the door, the lamp, the chair and so forth, but as directly responding to a "what for":

> What is first of all 'given' …is the 'for writing,' the 'for going in and out,' the 'for illuminating,' the 'for sitting.' That is, writing, going-in-and-out, sitting, and the like are what we are a priori involved with. What we know when we 'know our way around' and what we learn are these 'for-what's.[40]

It's clear here, unlike what some people take Heidegger to suggest in Being and Time, that this basic experience has no as-structure.[41] That is, when absorbed in coping, I can be described *objectively* as using the door *as* a door, but I'm not *experiencing* the door *as* a door. In coping at my best, I'm not experiencing *the door* at all but simply pressing into the possibility of going out. The important thing to realize is that, when we are pressing into possibilities, … there is no *experience* of an *entity* doing the soliciting; just the solicitation. Such solicitations disclose the world on the basis of which we sometimes do step back and perceive things *as* things.

But Agre's Heideggerian AI did not try to program this experiential aspect of being drawn in by an affordance. Rather, with his deictic representations, Agre *objectified* both the functions and their situational relevance for the agent. In Pengi, when a virtual ice cube defined by its function is close to the virtual player, a rule dictates the response, e.g. kick it. No skill is involved and no learning takes place.

So Agre had something right that I was missing -- the transparency of the ready-to-hand -- but he also fell short of being fully Heideggerian. For Heidegger, the ready-to-hand is not a fixed function, encountered in a predefined type of situation that triggers a predetermined response that either succeeds or fails. Rather, as we have begun to see and will soon see further, readiness-to-hand is experienced as a solicitation that calls forth a flexible response to the *significance* of the current situation – a response which is experienced as either improving one's situation or making it worse.

Although he proposed to program Heidegger's account of everyday routine activities, [42]Agre doesn't even try to account for how our experience feeds back and changes our sense of the significance of the next situation and what is relevant in *it*. By putting his virtual agent in a virtual world where all possibly relevance is determined beforehand, Agre can't account for how we learn to respond to new relevancies in our everyday routine activities, and so, like Brooks, he finessed rather than solved the frame problem. Thus, sadly, his Heideggerian AI turned out to be a dead end. Happily, however, Agre never claimed he was making progress towards building a human being.

V. Pseudo Heideggerian AI: Situated Cognition and the Embedded, Embodied, Extended Mind.

In Reconstructing the Cognitive World, Wheeler praises me for putting the confrontation between Cartesian and Heideggerian ontologies to an empirical test.

Wheeler claims, however, that, I only made negative predictions about the viability of GOFAI and Cognitive Science research programs. The time has come, he says, for a positive approach, and he claims that the emerging embodied-embedded paradigm in the field is a thoroughly Heideggerian one.

As if taking up from where Agre left off with his objectified version of the ready-to-hand, Wheeler tells us:

> [O]ur global project requires a defense of action-oriented representation. … [A]ction-oriented representation may be interpreted as the subagential reflection of online practical problem solving, as conceived by the Heideggerian phenomenologist. Embodied-embedded cognitive science is implicitly a Heideggerian venture. [43]

He further notes:

> As part of its promise, this nascent, Heideggerian paradigm would need to indicate that it might plausibly be able either to solve or to dissolve the frame problem.[44]

And he suggests:

> The good news for the reoriented Heideggerian is that the kind of evidence called for here may already exist, in the work of recent *embodied-embedded cognitive science*.[45]

He concludes:

> Dreyfus is right that the philosophical impasse between a Cartesian and a Heideggerian metaphysics can be resolved empirically via cognitive science. However, he looks for resolution in the wrong place. For it is not any alleged empirical failure on the part of orthodox cognitive science, but rather the concrete empirical success of a cognitive science with Heideggerian credentials, that, if sustained and deepened, would ultimately vindicate a Heideggerian position in cognitive theory.[46]

I agree it is time for a positive account of Heideggerian AI and of an underlying Heideggerian neuroscience, but I think Wheeler is looking in the wrong place. Merely in supposing that Heidegger is concerned with subagential *problem solving* and action oriented *representations,* Wheeler's project reflects not a step beyond Agre but a regression to pre-Brooks GOFAI. Heidegger, indeed, claims that that skillful coping is basic, but he is also clear that, at its best, coping doesn't involve representations or problem solving at all. [47]

Wheeler's cognitivist misreading of Heidegger leads to his overestimating the importance of Andy Clark's and David Chalmers' attempt to free us from the Cartesian idea that the mind is essentially inner by pointing out that in thinking we sometimes make use of external artifacts like pencil, paper, and computers.[48] Unfortunately, this argument for the extended mind preserves the Cartesian assumption that our basic way of relating to the world is by *thinking*, that is by using representations such as beliefs and memories be they in the mind or in notebooks in the world. In effect, while Brooks and Agre dispense with representations where coping is concerned, all Chalmers, Clark, and Wheeler give us as a supposedly radical new Heideggerian approach to the human way of being in the world is the observation that memories and beliefs are not necessarily inner entities and that, therefore, thinking bridges the distinction between inner and outer representations.[49]

When we solve problems, we do sometimes make use of representational equipment outside our bodies, but Heidegger's crucial insight is that *being-in-the-world* is more basic than *thinking* and solving problems*;* it is not representational at all. That is, when we are coping at our best, we are drawn in by affordances and respond directly to them, so that the distinction between us and our equipment--between inner and outer—vanishes.[50] As Heidegger sums it up:

> I *live* in the understanding of writing, illuminating, going-in-and-out, and the like. More precisely: as Dasein I am -- in speaking, going, and understanding -- an act of understanding dealing-with. My being in the world *is* nothing other than this already-operating-with-understanding in this mode of being.[51]

Heidegger's and Merleau-Ponty's understanding of embedded embodied coping, therefore, is not that the *mind* is sometimes *extended into the world* but rather that, in our most basic way of being, -- that is, as skillful copers, -- we are not minds at all but *one with the world*.

Heidegger sticks to the phenomenon, when he makes the strange-sounding claim that, in its most basic way of being, "Dasein is its world existingly."[52]

When you stop thinking that mind is what characterizes us most basically but, rather, that most basically we are absorbed copers, the inner/outer distinction becomes problematic. There's no easily askable question about where the absorbed coping is in me or in the world. Thus, for a Heideggerian all forms of *cognitivist* externalism presuppose a more basic *existentialist* externalism where even to speak of "externalism" is misleading since such talk presupposes a contrast with the internal. Compared to this genuinely Heideggerian view, extended-mind externalism is contrived, trivial, and irrelevant.

VI. What Motivates embedded/embodied coping?

But why is Dasein called to cope at all? According to Heidegger, we are constantly solicited to improve our familiarity with the world. Five years before the publication of Being and Time he wrote:

> Caring takes the form of a looking around and seeing, and as this circumspective caring it is at the same time … concerned about developing its circumspection, that is, about *securing and expanding* its *familiarity* with the objects of its dealings. [53]

This pragmatic perspective is developed by Merleau-Ponty, and by Samuel Todes.[54] These heirs to Heidegger's account of familiarly and coping describe how an organism, animal or human, interacts with the meaningless physical universe in such as way as to experience it as an environment organized in terms of that organism's need to find its way around. All such coping beings are motivated to get a more and more secure sense of the specific objects of their dealings. In our case, according to Merleau-Ponty:

> My body is geared into the world when my perception presents me with a spectacle as varied and as clearly articulated as possible, and when my motor intentions, as they unfold, receive the responses they anticipate [*attendent*, not expect] from the world.[55]

To take Merleau-Ponty's example:

> For each object, as for each picture in an art gallery, there is an optimum distance from which it requires to be seen, a direction viewed from which it vouchsafes most of itself: at a shorter or greater distance we have merely a perception blurred through excess or deficiency. We therefore tend towards the maximum of visibility, [as if seeking] a better focus with a microscope. [56]

In short, in our skilled activity we are drawn to move so as to achieve a better and better grip on our situation. For this movement towards maximal grip to take place, one doesn't need a mental representation of one's goal nor any subagential problem solving, as would a GOFAI robot. Rather, acting is experienced as a steady flow of skillful activity in response to one's sense of the situation. Part of that experience is a sense that when one's situation deviates from some optimal body-environment gestalt, one's activity takes one closer to that optimum and thereby relieves the "tension" of the deviation. One does not need to know what that optimum is in order to move towards it. One's body is simply solicited by the situation [the gradient of the situation's reward] to lower the tension. Minimum tension is correlated with achieving an optimal grip. As Merleau-Ponty puts it: "Our body is not an object for an 'I think', it is a grouping of lived-through meanings that moves towards its equilibrium."[57] [Equilibrium being Merleau-Ponty's name for zero gradient.]

VII. Modeling Situated Coping as a Dynamical System

Describing the phenomenon of everyday coping as being "geared into" the world and moving towards "equilibrium" suggests a *dynamic* relation between the coper and the environment. Timothy van Gelder calls this dynamic relation *coupling*. He explains the importance of coupling as follows:

> The fundamental mode of interaction with the environment is not to represent it, or even to exchange inputs and outputs with it; rather, the relation is better understood via the technical notion of coupling. ...
> The post-Cartesian agent manages to cope with the world without necessarily representing it. A dynamical approach suggests how this might be possible by showing how the internal operation of a system interacting

> with an external world can be so subtle and complex as to *defy* description in representational terms -- how, in other words, cognition can *transcend* representation. [58]

Van Gelder shares with Brooks the idea that thought is grounded in a more basic relation of agent and world. As van Gelder puts it:

> Cognition can, in sophisticated cases, [such a breakdown, problem solving and abstract thought] involve representation and sequential processing; but such phenomena are best understood as emerging from [i.e. requiring] a dynamical substrate, rather than as constituting the basic level of cognitive performance.[59]

This dynamical substrate is precisely the skillful coping first described by Heidegger and worked out in detail by Todes and Merleau-Ponty.

Van Gelder importantly contrasts the rich interactive temporality of real-time on-line coupling of coper and world with the austere step by step temporality of thought. Wheeler helpfully explains:

> [W]hilst the computational architectures proposed within computational cognitive science require that inner events happen in the right order, and (in theory) fast enough to get a job done, there are, in general, no constraints on how long each operation within the overall cognitive process takes, or on how long the gaps between the individual operations are. Moreover, the transition events that characterize those inner operations are not related in any systematic way to the real-time dynamics of either neural biochemical processes, non-neural bodily events, or environmental phenomena (dynamics which surely involve rates and rhythms).[60]

Computation is thus paradigmatically austere:

> Turing machine computing is digital, deterministic, discrete, effective (in the technical sense that behavior is always the result of an algorithmically specified finite number of operations), and temporally austere (in that time is reduced to mere sequence).[61]

Ironically, Wheeler's highlighting the contrast between rich dynamic temporal coupling and austere computational temporality enables us to see clearly that his appeal to extended minds as a Heideggerian response to Cartesianism leaves out the essential temporal character of embodied embedding. Clarke's and Chalmer's examples of extended minds dealing with representations are clearly a case of computational austerity. Wheeler is aware of this possible objection to his backing both the *dynamical systems* model and the *extended mind* approach. He asks. "What about the apparent clash between continuous reciprocal causation and action orientated representations? On the face of it this clash is a worry for our emerging cognitive science."[62] But, instead of engaging with the incompatibility of these two opposed models of ground level intelligence — on the one hand, *computation* as in GOFAI, classical Cognitivism, and Agre-like action-orientated representations, and on the other, *dynamical* models as demonstrated by Brooks and described by van Gelder — Wheeler punts. He simply suggests that we must somehow combine these two approaches and that "this is the biggest of the many challenges that lie ahead."[63]

Wheeler's ambivalence concerning the role of computation undermines his overall approach. This is not a mere local squabble about details, although Wheeler clearly wishes it were.[64] It is, as Wheeler himself sees, the issue as to which approach is more basic – the computational or the dynamic. The Heideggerian claim is that action-oriented coping as long as it is involved (on-line, Wheeler would say) is not representational at all and does not involve any problem solving, and that all representational problem solving takes place off-line and presupposed this involved coping,[65] Showing in detail how the representational un-ready-to-hand and present-at-hand in all their forms are derivative from non-representational ready-to-hand coping is one of Heidegger's priority projects.

More broadly, a Heideggarian cognitive science would require working out an ontology, phenomenology, and brain model that denies a basic role to austere computational processing, and defends a dynamical model like Merleau-Ponty's and van Gelder's that gives a primordial place to equilibrium, and in general to rich coupling. Ultimately, we will have to choose which sort of AI and which sort of neuroscience to back, and so we are led to our final questions: could the brain as its most basic way of making sense of things instantiate a richly coupled dynamical system, and is there any evidence it actually does so? If so, could this sort of non-computational coupling be modeled on a digital computer to give us Heideggerian AI?

Intermission

VIII Walter Freeman's Heideggerian/Merleau-Pontian Neurodynamics

We have seen that our experience of the everyday world is organized in terms of significance and relevance and that this significance can't be constructed by giving meaning to brute facts -- both because we don't experience brute facts and, even if we did, no value predicate could do the job of giving them situational significance. Yet, all that the organism can receive as input is mere physical energy. How can such senseless physical stimulation be experienced directly as significant? If we can't answer this question, the phenomenological observation that the world is its own best representation, and that the significance we find in our world is constantly enriched by our experience in it, seems to suggest that the brain is what Dennett derisively calls "wonder tissue."

Fortunately, there is at least one model of how the brain could provide the causal basis for the intentional arc. Walter Freeman, a founding figure in neuroscience and the first to take seriously the idea of the brain as a nonlinear dynamical system, has worked out an account of how the brain of an active animal can find and augment significance in its world. On the basis of years of work on olfaction, vision, touch, and hearing in alert and moving rabbits, Freeman proposes a model of rabbit learning based on the coupling of the brain and the environment. To bring out the relevance of Freeman's account to our phenomenological investigation, I propose to map Freeman's neurodynamic model onto the phenomena we have already noted in the work of Merleau-Ponty.

1. Involved action/perception. [Merleau-Ponty's being-absorbed-in-the-world (être au monde) -- his version of Heidegger's in-der-welt-sein.]

The animal will sometimes sense a need to improve its current situation. When it does, an instinct or a learned skill is activated. Thus, according to Freeman's model, when hungry, frightened, etc., the rabbit sniffs around seeking food, runs toward a hiding place, or does whatever else prior experience has taught it is appropriate. The animal's neural connections are then changed on the basis of the quality of its resulting experience, that is, they are changed in a way that reflects the extent to which the result satisfied the animal's current need. This is not simple behaviorism, however, since, as we shall now see, the changes brought about by experience are global, not discrete.

2. Holism

The change is much more radical than adding a new mechanical response. The next time the rabbit is in a similar state of seeking and encounters a similar smell the entire olfactory bulb goes into a state of global chaotic activity. Freeman tell us:

> [E]xperiments show clearly that every neuron in the [olfactory] bulb participates in generating each olfactory perception. In other words, the salient information about the stimulus is carried in some distinctive pattern of *bulb wide activity*, not in a small subset of feature-detecting neurons that are excited only by, say, foxlike scents.[66]

Freeman later generalizes this principle to 'brain-wide activity' such that a perception involves and includes all of the sensory, motor and limbic systems.

3. Direct perception of significance

After each sniff, the rabbit's bulb exhibits a distribution of what neural modelers traditionally call energy states. The bulb then tends toward minimum energy the way a ball tends to roll towards the bottom of a container, no matter where it starts from within the container. Each possible minimal energy state is called an *attractor*. The brain states that tend towards a particular attractor are called that attractor's *basin of attraction*.

The rabbit's brain forms a new basin of attraction for each new significant input.[67] Thus, the significance of past experience is preserved in the set of basins of attraction. The set of basins of attraction that an animal has learned forms what is called an *attractor landscape*. According to Freeman:

> The state space of the cortex can therefore be said to comprise an attractor landscape with several adjoining basins of attraction, one for each class of learned stimuli.[68]

Freeman argues that each new attractor does not *represent,* say*,* a carrot, or the smell of carrot, or even what to do with a carrot. Rather, the brain's current state is the result of the sum of the animal's past experiences with carrots, and this state is directly coupled with or resonates to the affordance offered by the current carrot. What in the physical input is directly picked up and resonated to when the rabbit sniffs, then, is the affords-eating.[69] Freeman tells us "The macroscopic bulbar patterns [do] not relate to the stimulus directly but instead to the significance of the stimulus."[70]**[Stuart asks: Are there attractors for carrot, celery, etc. or just for affords eating, running way from etc ?]**

Freeman adds:

> These attractors and behaviors are constructions by brains, not merely readouts of fixed action patterns. No two replications are identical: like handwritten signatures, they are easily recognized but are never twice exactly the same.[71]

<u>4. The stimulus is not further processed or acted upon. [Merleau-Ponty: We normally have no experience of sense data.]</u>

Since on Freeman's account the attractors are coupled directly to the significance of the current input, the stimulus need not be processed into a representation of the current situation on the basis of which the brain then has to infer what to do. So, after selecting and activating a specific attractor and modifying it, the stimulus has no further job to perform. As Freeman explains:

> The new pattern is selected, not imposed, by the stimulus. It is determined by prior experience with this class of stimulus. The pattern expresses the nature of the class and its significance for the subject rather than the particular event. The identities of the particular neurons in the receptor class that are activated are irrelevant and are not retained[72] ... Having played its role in setting the initial conditions, the sense-dependent activity is washed away.[73]

<u>5. The perception/action loop.</u>

The movement towards the bottom of a particular *perceptual* basin of attraction is correlated with the perception of the *significance* of a particular scent. It then leads to the animal's direct motor response to the current affordance, depending on how well that motor response succeeded in the past. According to Freeman, the perceptual "recognition" of the instrumental significance[74] of the current scent places the animal's *motor system* into into an appropriate basin of attraction. **[Stuart asks. how?]** For example, if the carrot affords eating the rabbit is directly readied to eat the carrot, or perhaps readied to carry off the carrot depending on which attractor is currently activated. Freeman tells us:

> The same global states that embody the significance provide… the patterns that make choices between available options and that guide the motor systems into sequential movements of intentional behavior.[75]

The readiness can change with each further sniff or shift in the animals attention like switching from frame to frame in a movie film.

But the changing attractor states are not fast enough to guide the animal's moment-by-moment motor responses to the changing situation. For that, the brain needs to switch to another form of processing that is directly responsive to the sensory input. This other form of processing must guide the moment-by-moment muscle contractions that control the animal's movements. It must therefore take account of how things are going and either continue on a promising path, or, if the overall action is not going as well as anticipated, it must signal the attractor system to jump to another attractor so as to increase the animals sense of impending reward.[76] If the rabbit achieves what it is seeking, a report of its success is fed back to reset the sensitivity of the olfactory bulb. And the cycle is repeated.

6. Optimal grip.

The animal's movements are presumably experienced by the animal as tending towards getting an optimal perceptual take on what is currently significant, and, where appropriate, an actual optimal bodily grip on it. Freeman sees his account of the brain dynamics underlying perception and action as structurally isomorphic with Merleau-Ponty's. He explains:

> Merleau-Ponty concludes that we are moved to action by disequilibrium between the self and the world. In dynamic terms, the disequilibrium ... puts the brain onto … a pathway through a chain of preferred states, which are learned basins of attraction. The penultimate result is not an

> equilibrium in the chemical sense, which is a dead state, but a descent for a time into the basin of an attractor, giving an awareness of closure.[77]
> **[Stuart says moving from one attractor to another requires an impossible discontinuous change in brain state. And also asks, what decides which attractor to move into?]**

Thus, according to Freeman, in governing action the brain normally moves from one basin of attraction to another descending into each basin for a time without coming to rest in any one basin. If so, Merleau-Ponty's talk of *reaching* equilibrium or maximal grip is misleading. But Merleau-Pontians should be happy to improve their phenomenological description on the basis of Freeman's model. Normally, the coper moves *towards* a maximal grip but, instead of coming to rest when the maximal grip is achieved, as in Merleau-Ponty's example of standing and observing a picture in a museum, the coupled coper, without coming to rest, is drawn to move on in response to the call of another affordance **[How do affordances call?]** that solicits her to take up the same task from another angle, or to turn to the next task that grows out of the current one.

<u>7. Experience feeds back into the look of the world. [Merleau-Ponty's intentional arc.]</u>

Freeman claims his read out from the rabbit's brain shows that each learning experience that is significant in a new way sets up a new attractor and rearranges all the other attractor basins in the landscape:

> I have observed that brain activity patterns are constantly dissolving, reforming and changing, particularly in relation to one another. When an animal learns to respond to a new odor, there is a shift in all other patterns, even if they are not directly involved with the learning. There are no fixed representations, as there are in [GOFAI] computers; there are only significances.[78]

Freeman adds:

> I conclude that the context dependence is an essential property of the cerebral memory system, in which each new experience must change all of the existing store by some small amount, in order that a new entry be incorporated and fully deployed in the existing body of experience. This

> property contrasts with memory stores in computers…in which each item is positioned by an address or a branch of a search tree. There, each item has a compartment, and new items don't change the old ones. Our data indicate that in brains the store has no boundaries or compartments.... Each new state transition … initiates the construction of a local pattern that impinges on and modifies the whole intentional structure.[79]

The whole constantly updated landscape of attractors is correlated with the agent's experience of the changing significance of things in the world.

The important point is that Freeman offers a model of learning which is not an associationist model according to which, as one learns, one adds more and more fixed connections, nor a cognitivist model based on off-line representations of objective facts about the world that enable inferences about which facts to expect next, and what they mean. Rather, Freeman's model instantiates a genuine intentional arc according to which there are no linear casual connections nor a fixed library of data, but where, each time a new significance is encountered, the whole perceptual world of the animal changes so that significance as directly displayed is contextual, global, and continually enriched.

8. Circular causality

Such systems are self-organizing. Freeman explains:

> Macroscopic ensembles exist in many materials, at many scales in space and time, ranging from…weather systems such as hurricanes and tornadoes, even to galaxies. In each case, the behavior of the microscopic elements or particles is constrained by the embedding ensemble, and microscopic behavior cannot be understood except with reference to the macroscopic patterns of activity…[80]

Thus, the cortical field controls the neurons that create the field. In Freeman's terms, in this sort of circular causality the overall activity "enslaves" the elements. As he emphasizes:

> Having attained through dendritic and axonal growth a certain density of anatomical connections, the neurons cease to act individually and start participating as part of a group, to which each contributes and from which

each accepts direction….The activity level is now determined by the population, not by the individuals. This is the first building block of neurodynamics. [81]

Given the way the whole brain can be tuned by past experience to influence individual neuron activity, Freeman can claim:

Measurements of the electrical activity of brains show that dynamical states of Neuroactivity emerge like vortices in a weather system, triggered by physical energies impinging onto sensory receptors. ... These dynamical states determine the structures of intentional actions. [82]

Merleau-Ponty seems to anticipate Freeman's neurodynamics when he says:

It is necessary only to accept the fact that the physico-chemical actions of which the organism is in a certain manner composed, instead of unfolding in parallel and independent sequences, are constituted… in relatively stable "vortices."[83]

In its dynamic coupling with the environment the brain tends towards equilibrium but continually [discontinuously] switching from one attractor basin to another like successive frames in a movie. In Freeman's terms:

Neocortical dynamics progresses through time by continual changes in state that adapt the cortices to the changing environment.[84]

The discreteness of these global state transitions from one attractor basin to another makes it possible to model the brain's activity on a computer. Freeman notes that:

At macroscopic levels each perceptual pattern of Neuroactivity is discrete, because it is marked by state transitions when it is formed and ended. ... I conclude that brains don't use numbers as symbols, but they do use discrete events in time and space, so we can represent them …by numbers in order to model brain states with digital computers.[85]

That is, the computer can model the anticipaion of input as well as the series of discrete transitions from basin to basin they trigger in the brain, thereby modeling how, on the basis of past experiences of success or failure, physical input acquires significance for the organism. When one actually programs such a model of the brain as a dynamic physical system, one has an explanation of how the brain does what Merleau-Ponty thinks the brain must be doing, and, since Merleau-Ponty is working out of Heidegger's ontology, one has developed Freeman's neurodynamics into Heideggerian AI.

Time will tell whether Freeman's Merleau-Pontian model is on the right track for explaining how the brain finds and feeds back significance into the meaningless physical universe. Only then would we find out if one could actually produce intelligent behavior by programming a model of the physical state transitions taking place in the brain. That would be the positive Heideggerian contribution to the Cognitive Sciences that Wheeler proposes to present in his book but which he fails to find. It would show how the emerging embodied-embedded approach, when fully understood, could, indeed, be the basis of a genuinely Heideggerian AI.

Meanwhile, the job of phenomenologists is to get clear concerning the phenomena that need to be explained. That includes an account of how *we,* unlike classical representational computer models, avoid the frame problem.

IX. How would Heideggerian AI dissolve the Frame Problem?

As we have seen, Wheeler rightly thinks that the simplest test of the viability of any proposed AI program is whether it can solve the frame problem. We've also seen that the two current supposedly Heideggerian approaches to AI avoid the frame problem. Brook's empiricist/behaviorist approach in which the environment directly causes responses avoids it by leaving out significance and learning altogether, while Agre's action-oriented approach, which includes only a small fixed set of possibly relevant responses, also avoids the problem.

Wheeler's approach, however, by introducing flexible action-oriented *representations*, like any representational approach has to face the frame problem head on. To see why, we need only slightly revise his statement of the problem (quoted earlier), substituting "representation" for "belief":

> [G]iven a dynamically changing world, how is a nonmagical system ….to retrieve and (if necessary) to revise, out of all the *representations* that it

> possesses, just those *representations* that are relevant in some particular context of action?[86]

Wheeler's frame problem, then, is to explain how his allegedly Heideggerian system can determine in some systematic way which of the action-oriented representations it contains or can generate are relevant in any current situation and keep track of how this relevance changes with changes in the situation.

Not surprisingly, the concluding chapter of the book where Wheeler returns to the frame problem to test his proposed Heideggerain AI, offers no solution or dissolution of the problem. Rather he asks us to "give some credence to [his] informed intuitions,"[87] which I take to be on the scent of Freeman's account of rabbit olfaction, that nonrepresentational causal coupling must play a crucial role. But I take issue with his conclusion that:

> in extreme cases the neural contribution will be nonrepresentational in character. In other cases, *representations* will be active partners alongside certain additional factors, but those representations will be action oriented in character, and so will realize the same content-sparse, action-specific, egocentric, context-dependent profile that Heideggerian phenomenology reveals to be distinctive of online *representational* states at the agential level.[88]

For Heidegger, *all* representational states are part of the problem.

Any attempt to solve the frame problem by giving any role to any sort of representational states even on-line ones has so far proved to be a dead end. So nonrepresentational action had better not be understood to be merely the "extreme case. " Rather, it must be, as Heidegger, Merleau-Ponty and Freeman see, our basic way of responding directly to relevance in the everyday world, so that the frame problem does not arise.

Heidegger and Merleau-Ponty argue that, thanks to our embodied coping and the intentional arc it makes possible, our skill in sensing and responding to relevant changes in the world is constantly improved. In coping in a particular context, say a classroom, we learn to ignore most of what is in the room, but, if it gets too warm, the windows solicit us to open them. We ignore the chalk dust in the corners and chalk marks on the desks but we attend to the chalk marks on the blackboard. We take for granted that what we write on the board doesn't affect the windows, even if we write, "open windows," and what we do with the windows doesn't affect what's on the board. And as we constantly refine this background know-how the things in the room and its layout become more and more familiar and take on more and more significance. In general, given our experience in the world, whenever there is a change in the current context we respond to it only if in the past it has turned out to be significant, and when we sense a significant change we treat everything else as unchanged except what our familiarity with the world suggests might also have changed and so needs to be checked out. Thus the frame problem does not arise.

But the frame problem reasserts itself when we need to change contexts. How do we understand how to get out of the present context and what to anticipate when we do? Merleau-Ponty has a suggestion. When speaking of one's attention being drawn by an object, Merleau-Ponty uses the term *summons* to describe the influence of a perceptual object on a perceiver.

> To see an object is either to have it on the fringe of the visual field and be able to concentrate on it, or else respond to this *summons* by actually concentrating on it.[89]

Thus, for example, as one faces the front of a house, one's body is already being *summoned* (not just *prepared*) to go around the house to get a better look at its back.[90]

Merleau-Ponty's treatment of what Husserl calls the *inner* horizon of the perceptual object e.g. its insides and back, applies equally to our experience of the object's *outer* horizon of other potential situations. As I cope, other tasks are *right now* present on the horizon of my experience summoning my attention as potentially (not merely possibly) relevant to the current situation. Likewise, my attention can be summoned by other potentially relevant situations already on the current situation's outer horizon.

If Freeman is right, this attraction of familiar-but-not-currently-fully-present aspects of what is currently ready-to-hand and of potentially relevant other familiar situations on the horizon might well be correlated with the fact that our brains are not simply in one attractor basin at a time but are influenced by other attractor basins in the same landscape, and by other attractor landscapes. **[How are they influenced?]**

According to Freeman, what makes us open to the horizonal influences of other attractors instead of our being stuck in the current attractor is that the whole system of attractor landscapes collapses and is rebuilt with each new rabbit sniff, or in our case, presumably with each shift in our attention. And once one correlates Freeman's neurodynamic account with Merleau-Ponty's description of the way the intentional arc feeds our past experience back into the way the world appears to us so that the world solicits from us appropriate responses, the problem of how we are summoned by what is relevant in our current situation, as well as other bordering situations, no longer seems insoluble.

But there is a generalization of the problem of relevance, and thus of the frame problem, that seems intractable. In What Computers Can't Do I gave as an example how, in placing a racing bet, we can usually restrict ourselves to such facts as the horse's age, jockey, past performance, and competition, but there are always other factors such as whether the horse is allergic to goldenrod or whether the jockey has just had a fight with the owner, which may in some cases be decisive. Human handicappers are capable of recognizing the relevance of such facts when they come across them.[91] But since anything in experience can be relevant to anything else, such an ability seems magical.

Jerry Fodor follows up on my pessimistic remark:

> "The problem," he tells us, "is to get the structure of an entire belief system to bear on individual occasions of belief fixation. We have, to put it bluntly, no computational formalisms that show us how to do this, and we have no idea how such formalisms might be developed. … If someone --a Dreyfus, for example-- were to ask us why we should even suppose that the digital computer is a plausible mechanism for the simulation of global cognitive processes, the answering silence would be deafening.[92]

However, once we give up computational Cognitivism, and see ourselves instead as basically coupled copers, we can see how the frame problem can be dissolved by an appeal to existential phenomenology and neurodynamics. In the light of how learning our way around in the world modifies our brain and so builds significance and relevance into the world so that relevance is directly experienced in the way tasks summon us, even the general problem raised by the fact that anything in our experience could in principle be related to anything else no longer seems a mystery.

X. Conclusion

It would be satisfying if we could now conclude that, with the help of Merleau-Ponty and Walter Freeman, we can fix what is wrong with current allegedly Heideggerian AI by making it more Heideggerian. There is, however, a big remaining problem. Merleau-Ponty's and Freeman's account of how we directly pick up significance and improve our sensitivity to relevance depends on our responding to what is significant for *us* given our needs, body size, ways of moving, and so forth, not to mention our personal and cultural self-interpretation. If we can't make our brain model responsive to the *significance* in the environment *as it shows up specifically for human beings*, the project of developing an embedded and embodied Heideggerian AI can't get off the ground.

Thus, to program Heideggerian AI, we would not only need a model of the brain functioning underlying coupled coping such as Freeman's but we would also need—and here's the rub—a model of *our particular way of being embedded and embodied s*uch that what we experience is significant for us in the particular way that it is. That is, we would have to include in our program a model of a body very much like ours with our needs, desires, pleasures, pains, ways of moving, cultural background, etc.[93]

So, according to the view I have been presenting, even if the Heideggerian/Merleau-Pontian approach to AI suggested by Freeman is ontologically sound in a way that GOFAI and the subsequent supposedly Heideggerian models proposed by Brooks, Agre, and Wheeler are not, a neurodynamic computer model would still have to be given a body and motivations like ours if things were to count as significant for it so that it could learn to act intelligently in our world. The idea of super-computers containing detailed models of human bodies and brains may seem to make sense in the wild imaginations of a Ray Kurzweil or Bill Joy, but they haven't a chance of being realized in the real world.

---

[1] This isn't just my impression. Philip Agre, a PhD's student at the AI Lab at this time, later wrote:

> I have heard expressed many versions of the propositions …that philosophy is a matter of mere thinking whereas technology is a matter of real doing, and that philosophy consequently can be understood only as deficient.

Philip E. Agre, Computation and Human Experience, (Cambridge: Cambridge University Press, 1997), 239.

[2] Marvin Minsky as quoted in a 1968 MGM press release for Stanley Kubrick's 2001: A Space Odyssey.

[3] Newell, A. and Simon, H.A., "Computer Science as Empirical Inquiry: Symbols and Search", Mind Design, John Haugeland, Edt. Cambridge, MA, MIT Press, 1988.

[4] Martin Heidegger, Being and Time, J. Macquarrie & E. Robinson, Trans., (New York: Harper & Row, 1962), 132, 133.

[5] Michael Wheeler, Reconstructing the Cognitive World: The Next Step, (Cambridge, MA: A Bradford Book, The MIT Press, 2005), 179.

[6] Edmund Husserl, *Experience and Judgment* (Evanston: Northwestern University Press, 1973), 38.

Roger Schank proposed what he called *scripts* such as a restaurant script, "A script, he wrote, "is a structure that describes appropriate sequences of events in a particular context. A script is made up of slots and requirements about what can fill those slots. The structure is an interconnected whole, and what is in one slot affects what can be in another. A script is a predetermined, stereotyped sequence of actions that defines a well-known situation." R.C. Schank and R.P. Abelson, Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures (Hillsdale, NJ: Lawrence Erlbaum, 1977) 41. Quoted in: Views into the Chinese Room: New Essays on Searle and Artificial Intelligence, John Preston and Mark Bishop, Eds, (Oxford: Clarendon Press, 2002).

[7]After I published, What Computers Can't Do in l972 and pointed out this difficulty among many others, my MIT computer colleagues, rather than facing my criticism, tried to keep me from getting tenure on the grounds that my affiliation with MIT would give

undeserved credibility to my "fallacies," and so would prevent the AI Lab from continuing to receive research grants from the Defense Department. The AI researchers were right to worry. I was considering hiring an actor to impersonate an officer from DARPA and to be seen having lunch with him at the MIT Faculty Club. (A plan cut short when Jerry Wiesner, the President of MIT, after consulting with Harvard and Russian computer scientists and himself reading my book, personally granted me tenure.) I did, however, later get called to Washington by DARPA to give my views, and the AI Lab did loose DARPA support during what has come to be called the AI Winter.

[8] Winograd, T. (1976). 'Artificial Intelligence and Language Comprehension,' in *Artificial Intelligence and Language Comprehension*, Washington, D.C.: National Institute of Education, 9.

[9] Wired Magazine, Issue 11:08, August 2003.

[10] Heidegger, Coping, and Cognitive Science, *Essays in Honor of Hubert L. Dreyfus*, Vol. 2, Mark Wrathall Ed, (Cambridge, MA: The MIT Press, 2000), iii.

[11] Michael Wheeler, Reconstructing the Cognitive World, 285.

[12] **Reference??**

[13] Terry Winograd, "Heidegger and he Design of Computer Systems," talk delivered at Applied Heidegger Conference, Berkeley, CA, Sept. 1989. Cited in What Computers *Still* Can't Do, Introduction to the MIT Press Edition, xxxi.

[14] Not everyone was pleased. One of the graduate students responsible for the invitation reported to me: "After it was announced that you were giving the talk, Marvin Minsky came into my office and shouted at me for 10 minutes or so for inviting you."

[15] What Computers *Still* Can't Do, *A Critique of Artificial Reason*. MIT Press, 1992, 265-266.

[16] Conclusion of footnote 11 in Edward Feigenbaum, "Artificial Intelligence: Themes in the second Decade," IFIP Congress '68, Final Supplement, p. J-13.

[17] What Computers *Still* Can't Do, 300.

[18] Rodney A. Brooks. "Intelligence without Representation," Mind Design, John Haugeland, Ed., The MIT Press, 1988, 416. (Brooks' paper was published in 1986) Haugeland explains Brooks' breakthrough using as an example Brooks' robot, Herbet:

> Brooks uses what he calls the "subsomption architecture", according to which systems are decomposed not in the familiar way by local functions or faculties, but rather by global *activities* or *tasks*. ... Thus, Herbert has one subsystem for detecting and avoiding obstacles in its path, another for wandering around, a third for finding distant soda cans and homing in on them, a fourth for noticing nearby soda cans and putting its hand around them, a fifth for detecting something between its fingers and closing them, and so on... fourteen in all. What's striking is that these are all complete input/output systems, more or less independent of each other. (John Haugeland, Having Thought: Essays in the Metaphysics of Mind, (Cambridge, MA: Harvard University Press, 1998), 218.)

[19] Ibid. 42.

[20] In fact he explicitly denies it saying:

> In some circles, much credence is given to Heidegger as one who understood the dynamics of existence. Our approach has certain similarities to work inspired by this German philosopher (for instance, Agre and Chapman 1987) but our work was not so inspired. It is based purely on engineering considerations. (Ibid., 415)

[21] Rodney A. Brooks, Flesh and Machines: How Robots Will Change Us**,** Vintage Books (2002), 168.

[22] "Can higher-level functions such as learning occur in these fixed topology networks of simple finite state machines?" he asks. ("Intelligence without Representation," Mind Design, 420.)

[23] Maurice Merleau-Ponty, Phenomenology of Perception, trans. C. Smith, Routledge & Kegan Paul, 1962, 136.

[24] See, Maurice Merleau-Ponty, The Structure of Behavior, A. L. Fisher, Trans, Boston: Beacon Press, 2nd edition 1966.

[25] Brooks. "Intelligence without Representation, 418. I**nsert Haugeland.**

[26] Rodney A. Brooks, "From earwigs to humans," Robotics and Autonomous Systems, vol. 20, 1997, 291.

[27] Daniel Dennett, "The Practical Requirements for Making a Conscious Robot," Philosophical Transactions of the Royal Society of London, A, v. 349, 1994, 133-146.

[28] Ibid. 133

[29] Although you couldn't tell it from the Cog web page. (**Give URL.**)

[30] Private communication. Oct. 26, 2005. (My italics.)

[31] Rodney A. Brooks, "From earwigs to humans,"301.

[32] Philip E. Agre, *The Dynamic Structure of Everyday Life*, MIT AI Technical Report 1085, October 1988, chapter 1, Section A1a, 9.

[33] Computation and Human Experience, 243**.** His ambitious goal was to "develop an alternative to the representational theory of intentionality, beginning with the phenomenological intuition that everyday routine activities are founded in habitual, embodied ways of interacting with people, places, and things in he world."

[34] Ibid, xi.

[35] Ibid. 332

[36] Ibid. 251. As Beth Preston sums it up in her paper, "Heidegger and Artificial Intelligence:" Philosophy and Phenomenological Research 53 (1), March 1993: 43-69**.**

> What results is a system that represents the world not as a set of objects with properties, but as current functions (what Heidegger called in-order-tos). Thus, to take a Heideggerian example, I experience a hammer I am using not as an object with properties but as in-order-to-drive-in-this-nail.

[37] Heidegger himself is unclear about the status of the ready-to-hand. When he is stressing the holism of equipmental relations, he thinks of the ready-to-hand as equipment, and of equipment as things like lamps, tables, doors, and rooms that have a place in a whole nexus of other equipment. Furthermore, he holds that breakdown reveals that these interdefined pieces of equipment are made of present-at-hand stuff that was there all along. (Being and Time, 97.) At one point Heidegger even goes so far as to include the ready-to-hand under the categories that characterize the present-at-hand:

> We call '*categories'* –characteristics of being for entities whose character is not that of Dasein." …"Any entity is either a "*who*" (existence) or a what (present-at-hand in the broadest sense.) **Page?**

---

[38] Being and Time, 405.

[39] Ibid. 99.

[40] Martin Heidegger, Logic: The Question of Truth, Trans. Thomas Sheehan manuscript. Gesamtausgabe, Band 21, 144. This is precisely what Brooks' animats cannot learn.

[41] Heidegger goes on immediately to contrast the total absorption of coping he has just described with the as-structure of thematic observation:

> Every act of having *things* in front of oneself and *perceiving them* is held within [the] disclosure of those *things*, a disclosure that things get from a *primary meaningfulness* in terms of the what-for. Every act of *having something in front of oneself and perceiving it* is, in and for itself, a 'having' s*omething as somethin*g.

To put it in terms of Being and Time, the as-structure of equipment goes all the way down in *the world,* but not in our *experience* of absorbed coping. It's bad phenomenology to read the self or the as-structure into our experience when we are coping at our best.

[42] **New insert from G**

[43] Michael Wheeler, Reconstructing the Cognitive World, 222-223.

[44] Ibid. 187.

[45] Ibid. 188.

[46] Ibid. 188-189.

[47] Merleau-Ponty says the same:

> [T]o move one's body is to aim at things through it; it is to allow oneself to respond to their call, which is made upon it independently of any representation. (Phenomenology of Perception, 139.)

[48] See, Clark, A. and Chalmers, D., "The Extended Mind," *Analysis* 58 (1): 7-19, 199.

[49] According to Heidegger, intentional content isn't in the mind, nor in some 3rd realm (as it is for Husserl), nor in the world; it isn't anywhere. It's a way of being-towards.

[50] As Heidegger puts it: "The self must forget itself if, lost in the world of equipment, it is to be able 'actually' to go to work and manipulate something." **Being and Time Page?**.

---

[51] Logic, 146. It's important to realize that when he introduces the term "understanding," Heidegger explains (with a little help from the translator) that he means a kind of know-how:

> In German we say that someone can *vorstehen* something—literally, stand in front of or ahead of it, that is, stand at its head, administer, manage, preside over it. This is equivalent to saying that he *versteht sich darauf*, understands in the sense of being skilled or expert at it, has the know-how of it. (Martin Heidegger, The Basic Problems of Phenomenology, A. Hofstadter, Trans. Bloomington: Indian University Press, 1982, 276.)

[52] Being and Time**,** 416**.** To make sense of this slogan, it's important to be clear that Heidegger distinguishes the human *world* from the physical *universe*.

[53] Martin Heidegger, *Phenomenological Interpretations in Connection with Aristotle,* in Supplements: From the Earliest Essays to *Being and Time* and Beyond, John Van Buren, Ed., State University of New York Press, 2002, 115. My italics.

This away of putting the source of significance covers both animals and people. By the time he published Being and Time, however, Heidegger was interested exclusively in the special kind of significance found in the world opened up by human beings who are defined by the stand they take on their own being. We might call this *meaning*. In this paper I'm putting the question of uniquely human meaning aside to concentrate on the sort of *significance* we share with animals.

[54] Todes goes beyond Merleau-Ponty in showing how our world-disclosing perceptual experience is structured by the actual structure of our bodies. Merleau-Ponty never tells us what our bodies are actually like and how their structure affects our experience. Todes notes that our body has a front/back and up/down orientation. It moves forward more easily than backward, and can successfully cope only with what is in front of it. He then describes how, in order to explore our surrounding world and orient ourselves in it, we have to be balanced within a vertical field that we do not produce, be effectively directed in a circumstantial field (facing one aspect of that field rather than another), and appropriately set to respond to the specific thing we are encountering within that field. For Todes, then, perceptual receptivity is an embodied, normative, skilled accomplishment, in response to

our need to orient ourselves in the world. (See, Samuel Todes, Body and World, Cambridge, MA: The MIT Press, 2001.)

[55] Merleau-Ponty, Phenomenology of Perception, 250. (Trans. Modified.)

[56] Ibid. 302

[57] Ibid, 153.

[58] Van Gelder, "Dynamics and Cognition", Mind Design II, John Haugeland, Ed., A Bradford Book, (Cambridge, MA: The MIT Press, 1997), 439, 448.

[59] Ibid.

[60] Michael Wheeler, *Change in the Rules: Computers, Dynamical Systems, and Searle*, in Views into the Chinese Room: New Essays on Searle and Artificial Intelligence, John Preston and Mark Bishop, Eds, (Oxford: Clarendon Press, 2002), 345.

[61] Ibid. 344, 345.

[62] Wheeler, Reconstructing the Cognitive World**, 2**80**.**

[63] Ibid.

[64] See Wheeler's footnote on the same subject declaring this confrontation a "minor spat" concerning "certain local nuances."(Ibid. 307.)

[65] I'm over simplifying here. Wheeler does note that Heidegger has an account of on-line, involved problem solving that Heidegger calls dealing with the un-ready-to-hand. The important points for Heidegger but not for Wheeler, however, are that (1) coping at its best deals directly with the ready-to-hand with no place for representations of any sort, and that (2) all un-ready-to-hand coping takes place on the background of an even more basic holistic coping which allows copers to orient themselves in the world. As we shall see, it is this basic coping, not any kind of problem solving, agential or subagential, that enables Heideggerian AI to avoid the frame problem.

[66] Walter J. Freeman, "The Physiology of Perception", Scientific American, 264: 78-85, 1991; and W. J. Freeman and K.A. Grajski, "Relation of olfactory EEG to behavior: Factor Analysis," Behavioral Neuroscience, 101: 766-777, 1987. **Page? My italics.)**

[67] In Freeman's neurodynamic model, the input to the rabbit's olfactory bulb modifies the bulb's neuron connections according to the Hebbian rule that neurons that fire together

wire together. Just how this Hebbian learning is translated into an attractor is not something Freeman claims to know in detail. He simply notes:

> The attractors are not shaped by the stimuli directly, but by previous experience with those stimuli, which includes preafferent signals and neuromodulators as well as sensory input. Together these modify the synaptic connectivity within the neuropil and thereby also the attractor landscape. [Walter Freeman, How Brains Make Up Their Minds, New York: Columbia University Press, 2000, 62]

[68] Walter Freeman, How Brains Make Up Their Minds, New York: Columbia University Press, 2000, 62. (Quotations from Freeman's books have been reviewed by him and sometimes modified to correspond to his latest vocabulary and way of thinking about the phenomenon.)

[69] **Should have a footnote on this being the brain activity presupposed by Gibson talk of resonating to affordances.**

[70]. Walter Freeman, Societies of Brains: A study in the neuroscience of love and hate, *The Spinoza Lectures, Amsterdam, Netherlands,* Hillsdale, N.J.: Lawrence Erlbaum Associates, Publisher, 1995, 59.

[71] Walter Freeman, How brains. 2000 , 62, 63. Merleau-Ponty is lead to a similar conclusion. **Ref. from Camillo.)**

[72], Walter Freeman, Societies of Brains, 66.

[73] Ibid. 67.

[74] See Sean Kelly, Logic of Motor Intenionality, **Ref.** Also, Corbin Collins describes the phenomenology of this motor intentionality and spells out the logical form of what he calls instrumental predicates. R**ef**

[75] Walter Freeman, How Brains Make Up Their Minds, 114.

[76] Freeman does not attempt to account for this direct control of moment-by-moment movements. To understand it we wold have to turn another form of nonrepresentational learning and skill called TDRL. See my paper with Stuart Dreyfus, **Ref**

[77] Ibid. 121

[78] Ibid. 22

[79] Walter Freeman, Societies of Brains. 99.

[80] Walter Freeman, How Brains Make Up Their Minds, 52

[81] Ibid. 53.

[82] Walter Freeman, Societies of Brains, 111.

[83] Maurice Merleau-Ponty, The Structure of Behavior, 153. Freeman says: "Merleau-Ponty's "vortices" correspond to the oscillations in brain potential that each attractor generates." (Private correspondence.)

[84] Walter Freeman, Societies of Brains, 100.

[85] Ibid. 105

[86] Wheeler, Reconstructing the Cognitive World, 179.

[87] Ibid. 279,

[88] Ibid. 276.

[89] Phenomenology of Perception, 67. My italics.

[90] Sean D. Kelly, " Seeing Things in Merleau-Ponty," in The Cambridge Companion to Merleau-Ponty**.**

[91] Hubert L. Dreyfus, What Computers Can't Do, (New York, NY: Harper and Row, 1997, 258.

[92] Jerry A. Fodor, The Modularity of Mind, Bradford/MIT Press,

[93] Dennett sees the "daunting" problem. He just doesn't see that it is a problem that we have no idea how to solve and which may well be insolvable. In a quotation I mentioned earlier he says further:

> Cog, …must have *goal-registrations* and *preference-functions* that map in rough isomorphism to human desires. This is so for many reasons, of course. Cog won't work at all unless it has its act together in a daunting number of different regards. It must somehow delight in learning, abhor error, strive for novelty, recognize progress. It must be vigilant in some regards, curious in others, and deeply unwilling to engage in self-destructive activity. While we are at it, we might as well try to make it crave human praise and company, and even exhibit a sense of humor.

(Consciousness in Human and Robot Minds For IIAS Symposium on Cognition, Computation and Consciousness, Kyoto, September 1-3, 1994, forthcoming in Ito, et al., eds., *Cognition, Computation and Consciousness*, OUP.)

We can, however, make some progress towards animal AI. Freeman claims his neurodynamic theory can be used to model lower organisms. In fact, he is actually using his brain model to program simulated robots. (See**:** Kozma R, Freeman WJ Basic principles of the KIV model and its application to the navigation problem. J Integrat. Neurosci 2: (2003 125-145.)

Freeman thinks that if he and his coworkers keep at it for a decade or so they might be able to model the body and brain of the salamander sufficiently to simulate its foraging and self-preservation capacitates. (Personal communication, 2/15/06)