Artificial Intelligence and the Body: Dreyfus, Bickhard, and the Future of AI

Daniel Susser

Abstract. For those who find Dreyfus's critique of AI compelling, the prospects for producing true artificial human intelligence are bleak. An important question thus becomes, what are the prospects for producing artificial non-human intelligence? Applying Dreyfus's work to this question is difficult, however, because his work is so thoroughly human-centered. Granting Dreyfus that the body is fundamental to intelligence, how are we to conceive of non-human bodies? In this paper, I argue that bringing Dreyfus's work into conversation with the work of Mark Bickhard offers a way of answering this question, and I try to suggest what doing so means for AI research.

Hubert Dreyfus's groundbreaking work in the philosophy of mind has demonstrated conclusively that the body is fundamental to all facets of intelligent life.[1] Thus Dreyfus has put to rest once and for all the formalist fantasy of a purely algorithmic, disembodied mind.[2] Furthermore, Dreyfus's constructive phenomenological work on skillful coping provides compelling reasons to believe that producing artificial human intelligence would effectively require replicating the human body,

Daniel Susser
Philosophy Department
Stony Brook University
e-mail: daniel.susser@stonybrook.edu

[1] This work first appeared in manuscript form in Dreyfus's (1972) *What Computers Can't Do: A Critique of Artificial Reason*, and was revised in 1979 and again in 1992, at which point it was re-issued by MIT Press under the title *What Computers Still Can't Do: A Critique of Artificial Reason*. As Dreyfus notes in the "Introduction to the Revised Edition," the book has remained largely intact since its first appearance, with only minor changes and new introductions with each new edition. All citations in this paper refer to the (1993) MIT Press second printing, as indicated in the list of references.

[2] Jerry Fodor, the arch-formalist, writes some ten years after Dreyfus first presented his argument, "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," quoted in Dreyfus, "Overcoming the Myth of the Mental: How Philosophers Can Profit from the Phenomenology of Everyday Expertise," *2005 APA Pacific Division Presidential Address*. Also see Bickhard and Terveen (1995), 42-44.

V.C. Müller (Ed.): Philosophy and Theory of Artificial Intelligence, SAPERE 5, pp. 277–287.
springerlink.com

socializing and enculturating it into everyday human life, and developing its capacities in more or less the same way human beings develop.[3] For those who take this account to be true, the question of artificial human intelligence has therefore ceased, more or less, to be a *philosophical* question, and has become instead a question for engineers. The question of whether or not artificial human intelligence is *possible* (and what the conditions of its possibility are) has become the question of whether or not it is *technologically feasible* to replicate the human body, embed the replica in human society, and so on.

Yet even in the wake of this analysis important philosophical questions remain unresolved. If one assumes, as I do, that producing artificial human intelligence (so conceived) is not feasible, a principal question becomes whether it is possible and what it might mean to produce some form of artificial *non-human* intelligence. After all, we find myriad forms of intelligence in the natural world: most people ascribe intelligence to cats and dogs; dolphins are without a doubt intelligent; and even certain birds and octopuses have demonstrated intelligent behavior. And while we can't necessarily understand such intelligent creatures fully, we can certainly understand some of them well enough to interact intelligently with them. Thus even if it's the case that we cannot produce artificial human intelligence, we might want to produce some other form of intelligence, some kind of intelligence which is neither human-like nor dog-like nor dolphin-like, but which is usefully or interestingly intelligent nonetheless.

If that is the case, however, we should need to ask whether and how Dreyfus's arguments about human intelligence pertain to such potential alternatives. Specifically, if the body is fundamental to all facets of intelligent life, it is presumably the case that any plausible form of non-human intelligence will have one. But what does that mean? What would an *artificial non-human body* look like? What is sufficient for constituting one? Indeed, what is it that all the various kinds of intelligent creatures found naturally have in common? What is common to human bodies and dog bodies and octopus bodies? Dreyfus's work doesn't make this entirely clear. For his analyses of skillful coping are derived phenomenologically, which is to say they are framed in terms of and articulated from the perspective of Dreyfus's own human subject position. He takes for granted that the intelligence under consideration is human intelligence and that the body is a human body.[4] In order to theorize artificial non-human intelligence while remaining true to Dreyfusian intuitions, it is therefore necessary to expand Dreyfus's analysis.

My goal in what follows is to show how we might begin to do that, and to offer some thoughts on what expanding the analysis means, theoretically and practically, for future artificial intelligence research. In order to do so, I attempt to bring Dreyfus's work into conversation with the work of Mark Bickhard, whose "interactivist" theory of cognition resembles Dreyfus's theory of skillful coping in crucial ways. Where they differ is that Bickhard's theory is oriented at a much higher level of generality than Dreyfus's. Instead of being framed in terms of human intelligence and human bodies, Bickhard's account is framed in terms of *physical*

[3] See Dreyfus's reply to Harry Collins in Dreyfus, "Responses," in *Heidegger, Coping, and Cognitive Science. Essays in Honor of Hubert L Dreyfus, vol. 2,* ed. Mark A. Wrathall and Jeff Malpas, 314-349. (Cambridge, Mass.: MIT Press, 2000), 345-6.

[4] Indeed, that it is a white, male, human body, etc.

systems generally. And thus it offers a way of extracting from Dreyfus's picture the basic features of bodies, common to all intelligent, embodied beings.

In the first two sections, I very briefly outline Dreyfus's theory of skillful coping and Bickhard's interactivist theory of cognition. Then I suggest how reading Dreyfus and Bickhard together offers a way of conceptualizing non-human intelligence that remains true to Dreyfusian intuitions, and what such a conception means for the future of AI.

1 Dreyfus on Intelligence and Skillful Coping

To understand Dreyfus's theory of skillful coping it is helpful to understand the critique against which it emerged. Now classic itself, Dreyfus's critique of classical (or computational, or "Good Old Fashioned") AI goes something like this: AI research is built upon two interrelated assumptions, both of which are false. First, that intelligence fundamentally is information processing (i.e., the manipulation of context-free symbols according to formal rules or algorithms); and second, that everything knowable about the world can be rendered in terms of discrete, independent representations. "In brief," Dreyfus writes in an early paper, "the belief in the possibility of AI, given present computers, is the belief that all that is essential to human intelligence can be formalized" (Dreyfus 1967: 1). But as Dreyfus points out, following Heidegger, Wittgenstein and others, there is a principled distinction between two aspects of human intelligence—knowing *that* (i.e., factual knowledge and reasoning about it) and knowing *how* (i.e., skills, behaviors, practices, etc.). On the computational view, *knowing that* is understood as fundamental, and all other intelligent skills and behaviors—everything from understanding language to recognizing faces—are taken simply to be "problems of complexity" (Dreyfus 1993: 55). That is to say, classical AI takes know-how to be derived from (and thus explainable in terms of) knowing-that. For Dreyfus, however, nothing could be further from the truth. Indeed, as he demonstrates, the computational view is not merely false, it is *backwards*. Our know-how is what fundamentally enables us to 'cope' with the world around us, not our formal reasoning. The former makes possible the latter.

The crux of Dreyfus's argument is that contrary to formalist desires, (1) meaning is inherently context-dependent, and (2) context-dependence in principle can't be formalized, because contexts are inherently indeterminate. Consider the following example, borrowed from Wittgenstein: walking down a country road, you come across a sign-post with an arrow on it. How do you know what the sign means? Should you follow the direction of the arrow or go in the opposite direction? Perhaps the sign is some sort of practical joke or was posted by someone who has a different understanding of arrows. What if the road curves? Should you deviate from it to continue in a straight line or follow along the curve? Is it significant that the sign is red? Or that a bird is flying in a particular direction overhead? It might be, if by local convention red signs indicate that one should follow the opposite direction of the arrow, or if one happened to know something about avian migration patterns. But then again it might not. The possibilities are endless.

This problem, known in linguistics and AI research as the "Frame Problem," is at bottom a matter of determining *relevance*. "Framing" something means

determining the appropriate context within which to understand it, and doing that amounts to determining what is and isn't relevant to its meaning. In the above example, understanding that the arrow on the sign means "go this way" necessarily involves knowing a few things about signs (and indeed, arrows). First of all, one must recognize it *as* a sign (instead of, say, a place to lock your bicycle). That way, one can determine that what is relevant to interpreting it has mostly to do with what is written on it (and not, say, its sturdiness, how well it is anchored in the ground, and so on). But even that initial recognition of the sign *as* a sign requires a larger context—namely, the context of being in the middle of a journey, and not at its end. Yet determining that context requires an even larger one, within which to understand the concept of a "journey." And so on, *ad infinitum*. Determining the appropriate context for understanding some phenomenon always requires appealing to another, larger context. Treating the problem formalistically therefore leads inevitably to regress.[5]

Since we, intelligent creatures, are nevertheless fully capable of doing it, of understanding things and the situations in which they arise, it seems then that we must do it in some other (non-formal, non-computational) way. Indeed, Dreyfus argues (following Heidegger), that the frame problem only arises in the first place because formalists have misunderstood the nature of intelligence. Formalists believe that intelligent creatures are *confronted with* situations, when in fact what normally happens is that we find ourselves *in* them.[6] On the former picture, an intelligence comprised of context-free facts and formal rules for manipulating them must reckon with a world of meaning fundamentally unlike itself—an unruly world, one which is contextual and indeterminate. It must either find or create a context within which to understand the phenomena at hand. The latter picture, however, suggests that the world and the intelligent actors in it are essentially of one piece. One need not find a context, for one is always already *in* one. This view suggests that we ought to understand the meaningful world as *our* world, as the world in which we are necessarily embedded, the world in which we live and act, and about which, sometimes, we think. The world understood in these terms is not a world comprised fundamentally of facts, but rather of tendencies, behaviors, practices, and skills. (It is comprised of facts, too, of course, but not fundamentally). This alternative to the formalist picture describes a world that is comprised, at bottom, of *know-how*. Furthermore, insofar as it is our world it is

[5] Dreyfus argues that this can be seen most clearly in modern formalist attempts at constructing psychological (or intentional) laws—in cognitive science, for example. The chief aim of cognitive scientists, according to Jerry Fodor, is to define computational mechanisms (i.e., formal rules or algorithms) that explain intentional laws (Fodor 1991: 20). All such laws, however, must contain ceteris paribus conditions. That is, they are necessarily 'non-strict', or apply 'everything else being equal' (21). Dreyfus argues that the ceteris paribus clause is essentially formal notation representing the background of human knowledge, and that "what 'everything else' and 'equal' means in any specific situation can never be fully spelled out without regress" (Dreyfus 1993: 57).

[6] Of course, we are sometimes confronted with a situation. Which is to say, we sometimes have to understand a situation from the perspective of an outside observer, such as when we watch a movie or the news. On the account I am presenting here, however, our capacity to understand such situations is parasitic upon a more fundamental form of understanding—namely, absorbed skillful coping.

one in which we fundamentally have a *stake*, a vested interest. The meaningful world is one about which intelligent actors have no choice but to be concerned. Sure, human beings may, at our relatively high level of intelligence, choose *how* to care about it, choose what to value more and what to value less. But insofar as we must act we are shaped and guided by our basic, inescapable interest in the way that activity relates us to and positions us within the world.[7]

On the whole, this background of know-how thus functions as a sort of global or ultimate context, shaping how we perceive the situations we find ourselves in, pre-reflectively selecting what is relevant for understanding them. Which is to say, our relationship to (and understanding of) the world is, at bottom, structured by our skills and skillful activity, and it is directed toward the satisfaction of those interests around which such skills develop in the first place. Put another way, this background enables us to *cope*. And it is here, for Dreyfus, that the body enters the picture. For in order to explain exactly how it is that this kind of coping or *skillful coping* works, he argues that we must look to our bodies.

At each moment and in every situation the body guides our sense of what is relevant, he claims, and it does so in three ways. The first has to do with brain architecture: "The possible responses to a given input must be constrained by […] this innate structure [which] accounts for phenomena such as the perceptual constants that are given from the start by the perceptual system as if they had always already been learned" (Dreyfus and Dreyfus 1999: 117). The brain, that is, acting as a transducer of sensory information intrinsically limits, by virtue of its physical architecture, the possible ways a situation can be perceived. We see only a certain part of the light spectrum, hear only certain wavelengths of sound, and the brain, though flexible, combines and interprets such sensory input in a relatively stable manner. The second way Dreyfus calls "body-dependent order of presentation." This describes how the physical structure of the body delimits the possible ways one might act in or interact with a given situation, and thus determines the range of possible ways one might understand it. "Things nearby that afford reaching," for example, "will be experienced early and often," etc. (Dreyfus and Dreyfus 1999: 118). The world is one in which we must act, and our bodies are such that only certain kinds of actions in certain situations are possible. Thus our understanding of the world is shaped each moment by the presence or absence of those various possibilities. Finally, the body guides our sense of what is relevant by aiming for what, following Merleau-Ponty, Dreyfus calls *maximum grip*—"the body's tendency to refine its discriminations and to respond to solicitations in such a way as to bring the current situation closer to the optimal gestalt that the skilled agent has learned to expect" (Dreyfus and Dreyfus 1999: 103). That is to say, the agent's overall sense of a situation implies an optimal relationship between the agent and the environment and "those input/output pairs will count as similar that move the organism towards maximum grip, which is itself a function of

[7] This anticipatory dimension of meaning is crucial and has to do with what Heidegger calls "care." See Dreyfus on Heidegger on care in Dreyfus, *Being-in-the-World: A Commentary on Heidegger's* Being and Time*, Division I* (Cambridge, Mass.: MIT Press, 1991), 238-45; Dreyfus, "Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian," *Philosophical Psychology* 20 (2007): 247-268.

body-structure" (Dreyfus and Dreyfus 1999: 118). Skillful activity is activity which can be more or less effective, successful or not. And the body is a barometer of this implicit normativity, tending naturally toward an agent-environment relationship in which its actions are best positioned to succeed.

In sum, on Dreyfus's account the body anchors us at the center of a perspective; it opens up a world. And it does so in three ways: first, by acting as a sensorial sieve, limiting at the outset what about the physical world can be *perceived*; second, by structuring the immediate environment around *possibilities for action*; and third, by pre-reflectively orienting movement toward the optimal relationship to (and understanding of) a given situation or some object in view. In other words, the body is what makes it possible to discover at any given moment that certain parts of the world are relevant to our interests or that they aren't, indeed to have interests at all. Our bodies embed us in a world of meaningful relations, make those relations matter to us, enable us to understand them (and ourselves in relation to them), and guide our activities in and through them.

2 Bickhard on Interaction and Recursive Self-maintenance

Mark Bickhard has developed a theory of cognition that is very much in the spirit of Dreyfus's account of skillful coping. Only instead of taking *human* intelligence as his object of analysis, as Dreyfus does, Bickhard aims to investigate intelligence construed more broadly. Bickhard thus articulates his theory in rather more general terms than Dreyfus does—namely, in terms not of the human body or human intelligence, but of the structures and functions of physical processes and systems. I take Bickhard's account to be so valuable for artificial intelligence research precisely for this reason, that it describes the actual requirements a physical system must meet in order to produce the capacity for some form of intelligence. It does not require that such systems are structured exactly like human beings; it merely requires that at some basic level humans and any such system have some organizational properties in common.[8]

[8] It should be noted at the outset that there appears to be a significant disparity between these two views. Dreyfus explicitly offers his approach as an alternative to *representationalist* views (i.e., those which take as a premise the notion that discrete, independent, content-bearing mental representations form the basic building blocks of intelligence.) Bickhard, on the other hand, specifically frames his interactivist theory as a new approach to theorizing representations. I believe that this apparent incompatibility is merely superficial. In the first place, Dreyfus does not deny the existence of mental representations; he merely denies that they are the fundamental components of cognition. That is to say, he rejects representational*ism*, not representations. Bickhard rejects representationalism too, only he calls it "encodingism" instead of "representationalism." And since Bickhard's whole project (with regard to representations) is to explain the processes that constitute them, he obviously agrees that they are not fundamental. Rather, on his interactivist account, certain kinds of complex physical processes produce representations. If my intuitions are correct, those complex processes are tantamount to Dreyfusian skills or know-how, and could presumably somehow produce representations for Dreyfus as well.

Bickhard's model centers around a type of open thermodynamic system known as a "dissipative structure." Such systems are characterized by the fact that they operate at far-from-thermodynamic-equilibrium conditions and *cease to exist* if such conditions are not maintained (Bickhard 2004: 11). Examples of such systems range from simple convection systems, such as those responsible for wind and rain, to the most complex systems in the universe—living organisms. Furthermore, within the class of far-from-equilibrium systems, a distinction can be made between those that require the explicit intervention of *another system* for maintaining its far-from-equilibrium conditions, and those which are able to some extent to maintain those conditions themselves (Bickhard 2004: 11). As an example of the former Bickhard points to a chemical bath, which requires that certain chemicals be constantly pumped into it in order to maintain its far-from-equilibrium state; an example of the latter is a candle flame, which "maintains above combustion threshold temperature; it melts wax so that it percolates up the wick; it vaporizes wax in the wick into fuel; [and] in standard atmospheric gravitational conditions, it induces convection, which brings in fresh oxygen and gets rid of waste" (Bickhard 2004: 11).

Bickhard refers to systems—such as the candle flame—which contribute to the preservation of their own far-from-equilibrium conditions as "self-maintaining systems". Any such system is, by definition, in constant *interaction* with its environment, because "self-maintenance is a(n emergent) property that is relative to a range of environments" (Bickhard 2004: 23). For instance, in the case of the candle flame, its self-maintaining processes will *fail* to preserve far-from-equilibrium conditions—whereby the system (flame) will cease to exist—if its environment changes in certain ways, such as there being no more wax or oxygen, etc. There are, however, more complex systems than candle flames, and some such systems can interact with their environments in more complex ways.

> The candle flame has no options, but other systems do. A bacterium, for example, might swim so long as it is swimming up a sugar gradient, but tumble if it finds itself swimming down a sugar gradient […] The swimming is self-maintaining so long as it is oriented toward higher sugar concentrations, but it is *not* self-maintaining if it is oriented toward lower sugar gradients. Conversely with tumbling. So, swimming is self-maintenant [sic] under some conditions and not under others, and the bacterium can detect the difference in the conditions and switch its activities accordingly; it can select between a pair of possible interactive processes that which would be appropriate for current (orientation) conditions (Bickhard 2004: 23-4).

In other words, the bacterium can (inter)actively maintain its very process of self-maintenance by distinguishing between variable environmental conditions—that is, by distinguishing between the presence of food (more sugar) and not-food (less sugar). It possesses interconnected subsystems, each of which can behave in different ways depending on the states of the other systems. This ability to (inter)actively detect what counts as the proper functioning of a system, given certain environmental conditions, is what Bickhard refers to as *recursive self-maintenance* (Bickhard 2004: 24). And it is this capacity that he suggests gives rise to cognition.

To see how this happens, consider the bacterium again. We saw, above, that in order to select whether to swim or tumble, it must be able to differentiate between environments that make one or the other behavior self-maintaining. Of course, we wouldn't say that when swimming it must, therefore, *know* that it is moving up a sugar gradient. How, then, can we explain what goes on when it detects environmental conditions, and the result of that detection causes it to behave (i.e., to interact further with the environment) some way rather than another? Bickhard's response comes in two parts: first, the state of some subsystem (e.g., a subsystem that detects sugar) "*implicitly define[s]* the class of environments that would yield that state if in fact encountered in an interaction" (Bickhard and Terveen 1995: 60); and second, some other subsystem (e.g., one which selects whether to swim or tumble) *functionally presupposes* that the environment is a certain way—based on the current state of the 'first' subsystem—and responds accordingly (Bickhard 2004: 25). The state of the first subsystem, that is, *implies* that certain environmental conditions obtain (and that others don't) in the same way that the mercury level of a thermometer reaching the notch marked "73 F" implies that it is seventy three degrees Fahrenheit (and that it is *not* forty three degrees Fahrenheit). The second subsystem then acts based on the discrimination made by the first. Bickhard calls this process, wherein one subsystem utilizes the state of another, *functional presupposition.* Thus on the interactive model one subsystem utilizes the state of another—the former functionally presupposes what is implied by the latter—to determine the type of behavior that will contribute to the maintenance of its own far-from-equilibrium conditions in a given situation. This complex process is, for Bickhard, the foundation of intelligent behavior.

While it is outside the scope of this paper to elaborate either Bickhard's or Dreyfus's picture more fully, I believe that we can already see a shared understanding of intelligence at work.[9] For both theorists, intelligence is a matter, more or less, of acting skillfully to satisfy one's needs and interests, and where doing so means interacting dynamically with the world in which one is fundamentally, inexorably embedded. Indeed, it seems to me that the process described above, wherein a physical system maintains its own existence conditions by successfully discriminating between healthy and toxic environments and by tending toward the former, *is* skillful coping in its simplest form, that this is a description of Dreyfus's concept of skillful coping at a higher level of generality. Furthermore, and to return to the question with which this essay began, I would like to suggest that Bickhard's characterization of *recursively self-maintaining physical systems* is as good a definition as any of what physically constitutes a *body*.

What is indispensable about Bickhard's view is that it glimpses these fundamental components of skillful coping in even their most primitive incarnations. And Bickhard does so not only by pointing metaphorically to the sort of "lower" cognitive functioning which humans share with non-human animals, as Dreyfus does, but by elaborating how such primitive intelligence works and how "higher

[9] I have argued at greater length for the parallel between Dreyfus's and Bickhard's conceptions of skillful coping and interactive cognition, respectively, in "Challenging the Binary: Toward an Ecological Theory of Intentionality," my 2007 philosophy honors thesis at The George Washington University.

level" intelligence might plausibly arise out of it. This is important because it encourages us to think about intelligence from the ground up, so to speak, rather than from the top down. It gives us a way of thinking about building artificial intelligence, instead of artificial *human* intelligence. That is, it suggests why and how we ought to think about building simple artificially intelligent systems, rather than attempting to reverse-engineer ourselves. In what remains, I will elaborate on this a bit, and suggest what I take it to mean, practically, for AI research.

3 Three Considerations for Future Research

Reading Dreyfus and Bickhard together leads to a generalized conception of the body as *an open thermodynamic system with the capacity to contribute to the maintenance of its own existence conditions by interacting skillfully with its environment.*[10] And while I am unable here to argue more fully and persuasively for this view, I would like to suggest that understanding bodies in this way brings certain important features of the relationship between intelligence and embodiment to the fore.

First, it indicates that bodies and intelligence are not distinct things. The claim that the body is fundamental to all facets of intelligent life is not merely the claim that bodies and intelligence are co-extensive, that wherever intelligence is found so too is there a body. Rather, it is the much stronger claim that bodies *are* intelligent. The more or less discrete physical systems we call bodies are just the sort of physical systems with the capacity to interact skillfully with their environments. The distinction between bodies and intelligence is an *analytical* distinction—it refers to two aspects of the same phenomenon (its physical properties and its skills or capacities).

This is a point which seems to me to have been lost on many of those who take Dreyfus's work very seriously. Thus one finds AI researchers attempting to strap humanoid robot "bodies" onto complex computers, or conversely, trying to capture the dynamics of embodiment in complex digital models.[11] In both cases, the body is understood as something that intelligence *requires*, a necessary feature which must be supplied or involved or made reference to, instead of being understood as what intelligence *is*. But the point that Dreyfus is making in his work is precisely that such a conception is misguided, that intelligence and the body are inseparable, that they are two sides of the same coin, that they develop together in the world, that intelligent creatures are intelligent because they are embodied, and

[10] It is worth noting that similar definitions have been put forward to describe *life*. And indeed, for many of the reasons outlined above, I would not be surprised if artificial intelligence and artificial life were developed simultaneously. Put another way, I think truly artificially intelligent systems will be difficult to distinguish from living ones.

[11] The former is evident in work such as Rodney Brooks and Daniel Dennett's "Cog" project, the latter in Walter Freeman's "neurodynamic modeling." For detailed accounts of both of these approaches, as well as a general survey of the state of the art in "Heideggerian AI," see Dreyfus, "Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian," op. cit.

as a result, that intelligence must be understood in terms of embodied activity in the world.[12] For AI researchers going forward, then, the first point to consider is that artificial intelligence and artificial embodiment must be developed in tandem. "Hardware" and "software" cannot be understood as fundamentally distinct. Instead, the very organizational structure of physical systems must be designed to produce intelligent behavior. In order to develop truly intelligent systems, we must design physical systems whose raisons d'être are to cope with their environments.[13]

Second, as the above suggests, this requires thinking about artificial intelligence on a much smaller scale. Instead of aiming for complex intelligent systems, researchers should try to build physical systems with small skill sets, but which also have the flexibility to adapt and learn. In this way, complexity can emerge out of simple intelligent systems.[14] Genetic programming and "generative AI" seem to me to be promising avenues of research that approach AI in just this way.[15] So too is the "enactive approach," developed most prominently by Evan Thompson and Francisco Varela. That approach offers a conception of the relationship between embodiment and intelligence similar to the one advocated here and has produced significant work in philosophy and cognitive science.[16]

Finally, understanding intelligence and the body in the way I've described suggests that AI researchers ought to be thinking not only about *how* intelligent creatures are intelligent, but also about *why* they are intelligent. As Dreyfus has shown, following Heidegger, meaningfulness and intelligence arise in the pursuit of interests, in relation to a world in which one is inexorably embedded—a world about which one has no choice but to care. Bickhard's conception of recursively self-maintaining systems brings this notion into even sharper relief: building beings that understand the world—in whatever way they do—and that are able, therefore, to behave intelligently in the world, means building beings that *need* to be intelligent in order to successfully function. This constitutive need for intelligence is crucial to understanding intelligence as such.[17] The body is what produces this need, what anchors intelligent creatures in the world, what *invests* us in it, what makes the world relevant and significant to us, what makes it such that we *have to cope*. Bodies, in a word, are why intelligence matters.

---

[12] This is another way of describing what Merleau-Ponty calls "the flesh," a notion which has undoubtedly shaped Dreyfus's thinking.

[13] For a marvelous discussion of both theoretical and experimental work related to this idea, see Slawomir Nasuto and Mark Bishop's "Of (Zombie) Mice and Animats" in this volume.

[14] For a helpful discussion of how this kind of emergence works, see Bickhard, "Emergence," in *Downward Causation*, ed. P. B. Andersen, C. Emmeche, N. O. Finnemann, P. V. Christiansen, 322-348. (Aarhus, Denmark: University of Aarhus Press, 2000).

[15] See Tijn van der Zant, *Generative AI: A Neo-Cybernetic Analysis* (Groningen: University Library Groningen).

[16] For an overview of the enactive approach, see Thompson's *Mind in Life: Biology, Phenomenology, and the Sciences of Mind* (Cambridge: The Belknap Press of Harvard University Press, 2007).

[17] See Nasuto and Bishop's paper (op cit.) for more on the constitutive need for and drive toward intelligence.

References

Bickhard, M.: Part II: Applications of Process-Based Theories: Process and Emergence: Normative Function and Representation. Axiomathes 14(1), 121–155 (2004)

Bickhard, M.: Emergence. In: Andersen, P.B., Emmeche, C., Finnemann, N.O., Christiansen, P.V. (eds.) Downward Causation, pp. 322–348. University of Aarhus Press, Aarhus (2000)

Bickhard, M.H., Terveen, L.: Foundational issues in artificial intelligence and cognitive science: impasse and solution. Elsevier, Amsterdam (1995)

Dreyfus, H.L.: Why Computers Must Have Bodies in Order to Be Intelligent. The Review of Metaphysics 21(1), 13–32 (1967)

Dreyfus, H.L.: Being-in-the-world: a commentary on Heidegger's Being and time, division I. MIT Press, Cambridge (1991)

Dreyfus, H.L.: What computers still can't do: a critique of artificial reason. MIT Pr., Cambridge (1993)

Dreyfus, H.L.: Responses. In: Wrathall, M.A., Malpas, J. (eds.) Heidegger, Coping, and Cognitive Science. Essays in Honor of Hubert L Dreyfus, vol. 2, pp. 314–349. MIT Press, Cambridge (2000)

Dreyfus, H.L.: Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian. Philosophical Psychology 20(2), 247–268 (2007)

Dreyfus, H.L.: Overcoming the Myth of the Mental: How Philosophers Can Profit from the Phenomenology of Everyday Expertise. APA Pacific Division Presidential Address (2005)

Dreyfus, H.L., Dreyfus, S.E.: The Challenge of Merleau-Ponty's Phenomenology of Embodiment for Cognitive Science. In: Weiss, G., Haber, H.F. (eds.) Perspectives on Embodiment: The Intersection of Nature and Culture. Routledge, New York (1999)

Fodor, J.: You Can Fool Some of the People All of the Time, Everything Else Being Equal: Hedged Laws and Psychological Explanations. Mind 100(1), 19–34 (1991)

van der Zant, T.: Generative AI: A Neo-Cybernetic Analysis. University Library Groningen, Groningen (2010)