



## HEIDEGGER AMONG THE ROBOTS

Alistair Welchman (University of Texas at San Antonio)

*Cognitive science and artificial intelligence have undergone some revolutionary changes in the past two decades. From an emphasis on disembodied cognitive functions like chess and logic, they now foreground the embodied and environmentally embedded nature of intelligent action. Some—both philosophers of cognitive science and practitioners—have sought to explain this shift in terms of a Heideggerian critique of the residually Cartesian assumptions of the traditional picture of disembodied cognition. I support the opening up of new areas of research practice formally closed off by tacit and unjustified theoretical presuppositions, but argue that these changes are and have been warranted by biological and information-theoretic concerns and not phenomenological ones derived from Heidegger's thought.*

### I

Cognitive science and artificial intelligence have undergone some revolutionary changes in the past two decades, changes that have corresponded with a general renewal of interest in both disciplines after the doldrums of the 1980s.<sup>1</sup> Traditionally these disciplines had been concerned with cognitive abilities like chess, first-order logic, highly artificial navigation tasks, or natural language processing. These tasks were abstracted from their neural implementation, and usually both inputs to and outputs from the tasks were couched in symbolic form. This conception of intelligence has several consequences. First, peripheral bodily perceptual and motor systems are excluded from cognition proper and reduced to transducers in and out of symbolic form. Second, the environment external to the transducers can be seen only as a source of possible information and

---

<sup>1</sup> Insofar as cognitive science is not just the same thing as cognitive psychology, it takes its cue from artificial intelligence and has traditionally shared assumptions with it (*e.g.*, the assumption that the mind is or can be modelled as a computer). In what follows I will usually refer simply to artificial intelligence for brevity.



arena for motor action. Third, this passive picture of cognition effectively forecloses the possibility that motor engagement might play an active role in cognition. Fourth, the requirement for symbolic inputs and outputs to the central system implies that the task of the peripheral systems is to encode inputs in a representational form ready for presentation to the symbolic processing unit (and conversely for motor outputs). These peripheral systems are often neglected on the grounds that transduction was essentially a simple engineering issue. But these supposedly easy engineering tasks turn out not to be so easy at all: linguistic encoding and perceptual modelling for instance are still—more than a half century after the artificial intelligence pioneers—hardly solved problems.

Since the early 1990s each of the assumptions of traditional artificial intelligence has been challenged: it has been claimed that cognition—or more broadly, intelligent action<sup>2</sup>—is fundamentally embodied, environmentally situated, and enactive. The first claim challenges the assumption that the essence of cognition can be located in realization-independent algorithms in which the body serves only as the source of inputs and target of outputs. It maintains instead that the nature and structure of the body make an ineliminable contribution to cognition.<sup>3</sup> The second claim rejects the view that the environment is primarily a mere source of information and instead regards cognition as emerging from co-operation between embodied agent and surrounding environment. Clark and Chalmers<sup>4</sup> present the most vivid form of this claim in their “extended mind” thesis according to which the mind is literally comprised of neurological and environmental components. Their classic example is the notebook used by an Alzheimer’s sufferer as a short-term memory. The third claim contests the traditional picture in which motor action is simply the output of the cognitive system, claiming instead that the activity of the intelligent agent is an essential part of cognition.<sup>5</sup> Lastly, as the source of intelligent action is spread over body and

---

<sup>2</sup> This term covers cases—of particular interest to the revolutionaries—in which biological or robotic activity is successful (by some measure) but not necessarily mediated by anything like knowledge that would raise it to the level of propositionally encoded cognition.

<sup>3</sup> See Shaun Gallagher, *How the Body Shapes the Mind* (Oxford: Oxford University Press, 2005).

<sup>4</sup> Andy Clark and David J. Chalmers, “The Extended Mind,” *Analysis*, vol. 58, no. 1 (1998), 7–19.

<sup>5</sup> See Francisco J. Varela, Evan Thompson and Eleanor Rosch, *The Embodied Mind* (Cambridge: MIT Press, 1992) and Alva Noë, *Action in Perception* (Cambridge: MIT Press, 2004).

world, the need for a representational interface between core and periphery becomes less pressing.

Robotics has formed a privileged venue for this reorientation in artificial intelligence because autonomous robots have to deal with their corporeal embodiment, their environmental situatedness, and the results of their own actions in ways that chess programs do not.<sup>6</sup> The problems faced by the early research robot SHAKEY were emblematic here: it took up to eight hours to perform simple tasks like picking up a block and taking it to another room despite inhabiting a mercilessly simplified environment.<sup>7</sup> The contrast with Brooks' agile, effortlessly mobile insectoid robots could not be clearer.<sup>8</sup>

Philosophically, it has become almost a commonplace to diagnose the problems of traditional artificial intelligence as the result of unconscious adherence to a Cartesian picture of the relation of intelligence to body, world, and action.<sup>9</sup> The simplest form of this diagnosis is that the explicitly dualist picture presented by Cartesian substance dualism has been prematurely naturalized. "Prematurely" is meant to indicate that certain structural features of the dualist (*i.e.*, non-naturalist) picture are unconsciously retained in the putatively naturalized version and that these features represent unsupported constraints on the formation of naturalistic hypotheses. For instance, where Descartes treated the mind as a separate substance from the extended matter comprising the body and world, traditional artificial intelligence thinks of both as material, but still splits mind from body in explanatory terms because the preferred vocabulary for explaining world and body is that of physics, whereas the mind should be explained in information-processing terms. Thus a new split between the intelligent and the physical runs along the same fault line

---

<sup>6</sup> See Rodney A. Brooks, "Intelligence without Reason," *MIT AI Lab Memo*, no. 1293 (1991), and "Intelligence without Representation," *Artificial Intelligence*, vol. 47, no. 1-3 (1991), 139-59.

<sup>7</sup> See Rodney A. Brooks, *Flesh and Machines: How Robots Will Change Us* (New York: Pantheon, 2002), 22f.

<sup>8</sup> *Ibid.*, Chapter 3.

<sup>9</sup> See Hubert L. Dreyfus, *Being-in-the-World: A Commentary on Heidegger's Being and Time Division 1* (Cambridge: MIT Press, 1991), Chapter 6; Tim Van Gelder, "What Might Cognition Be, if Not Computation?," *Journal of Philosophy*, vol. 92, no. 7 (1995), 345-81, here 380; Michael Wheeler, *Reconstructing the Cognitive World: The Next Step* (Cambridge: MIT Press, 2005); Shaun Gallagher and Dan Zahavi, *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science* (London and New York: Routledge, 2008), 135; Diane Proudfoot, "On Wittgenstein on Cognitive Science," *Philosophy*, vol. 72 (1997), 189-217. Hereafter Wheeler's text will be referred to parenthetically in the text as CW.

as Descartes' original ontological split, but now it is framed in explanatory rather than ontological terms.

Such an explanatory divide has, according to this story, effects similar to Descartes' original ontological divide: intelligence is abstracted from corporeal implementation, it requires transducers to filter information from the world and present it for intelligent processing (think: pineal gland), it treats the environment as at best a source of information to be represented (or conversely as an arena for post-processing intelligent action), and it regards action merely as the output of the system—at least on the current iteration. Thus the cluster of tacitly maintained assumptions comprising the specific structure of traditional artificial intelligence may itself be understood as arising from the tenacious grip of the picture of mind and world put forward by Descartes. An explanatory divide running parallel to Descartes' ontological divide makes clear why traditional artificial intelligence was so focussed on the manipulation of symbolic representations: representation, as for Descartes, is the mind's only way of connecting with the non-mental outside.

The idea that traditional artificial intelligence assumed a still recognizably Cartesian conception of intelligence suggests that the radical reorientation of artificial intelligence should in part take the form of a critique of Descartes. A properly conceived and philosophically naturalistic account of intelligence needs to do more than replace "mental substance" with "information processor" in the picture presented by Descartes' philosophy. Rather it demands a critique that will undertake a radical reconceptualization of intelligence before an appropriate naturalization can be successfully undertaken. Since there have already been a large number of philosophical critiques of Cartesian substance dualism, it may be that contemporary artificial intelligence can gain philosophical support from applying these to the explanatory Cartesianism of traditional artificial intelligence.<sup>10</sup>

In this paper I will evaluate the claim that Martin Heidegger's phenomenological critique of Descartes provides the philosophical resources for understanding the reorientation of contemporary artificial intelligence.<sup>11</sup> Since, as I have argued, robotics has a privi-

---

<sup>10</sup> See J. Petitot, F. Varela, B. Pachoud and J. M. Roy, eds., *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science* (Stanford: Stanford University Press, 1999).

<sup>11</sup> See Beth Preston, "Heidegger and Artificial Intelligence," *Philosophy and Phenomenological Research*, vol. 53, no. 1 (1993), 43–69; Wheeler, *Reconstructing the Cognitive World*; and Dreyfus, *Being-in-the-World*, Chapter 6.

leged place in the shift from traditional to contemporary artificial intelligence, this places Heidegger squarely among the robots. The question I want to answer is: should he be there?

In §II, I will outline the structure and motivation of Heidegger's argument in *Being and Time* in order to highlight some of the provisional convergences between Heidegger's work and that of contemporary artificial intelligence. In §III, I will address the obvious objection that this understanding of Heidegger is remote from Heidegger's actual thought: the purpose of the critique of Descartes from the point of view of artificial intelligence is to purge residual dualism so as to be able to clear the way for a proper naturalization of cognition and intelligent action. But Heidegger is himself fairly uncontroversially not a proponent of any philosophical naturalization of the mind. Then, in §IV, I will suggest how it is nevertheless possible to appropriate Heidegger's phenomenology in a way that is at least compatible with naturalism about cognition, despite Heidegger's probable hostility to such a move. Lastly, in §V, I will evaluate this move.

## II

In this section I give an account of the *prima facie* convergence between Heidegger's philosophical critique of Descartes and the parallel critique within contemporary artificial intelligence of the residual Cartesianism of traditional artificial intelligence.

The *locus classicus* for Heidegger's critique of Descartes is Division One of his 1927 book *Being and Time*.<sup>12</sup> As with all of his work, the philosophically primary question he addresses there is that of the meaning of (the question of) being in general. Here it is not necessary to deal directly with that question because in this text Heidegger argues that there is a preliminary set of questions that

---

<sup>12</sup> Martin Heidegger, *Sein und Zeit*, 18. Auflage (Tübingen: Max Niemeyer, 1927), (tr.) J. Macquarrie and E. Robinson as *Being and Time* (New York: Harper and Row, 1962). Hereafter referred to parenthetically in the text as SZ. References to Heidegger's text are to the German pagination and/or section numbers, both retained in the English version. Translations from Heidegger are my own, but usually follow Macquarrie and Robinson, with the exception of the crucial technical terms *das Vorhandene* and *das Zuhandene* (see below), which are translated as the "at-hand" and the "to-hand" respectively. In addition, I have rendered *das Seiende* as "being" rather than "entity" and eliminated capitalization for "Being" when it is translating *das Sein* since no ambiguities are introduced by doing so in this paper.

must be answered concerning the nature or being of *Dasein*, the being that each of us is.<sup>13</sup> The justification for the methodological priority of an investigation of the being of *Dasein* is that *Dasein* is distinguished from other beings by the fact that “in its being this being is an *issue* for it,” *i.e.*, *Dasein* is the only being that concerns itself with its own being. (SZ, 12) Heidegger reserves the term “existence” exclusively to refer to the being of *Dasein* so that Division One of *Being and Time* takes the form of an “existential analytic,” *i.e.*, an analysis of existence, *i.e.*, of the being of *Dasein*. A major target of this phenomenological analysis is what Heidegger regards as the mistaken account of the being of *Dasein* given by Descartes. (SZ, §6, §§19–21)

Some features of the critique of residual Cartesianism in traditional artificial intelligence are conspicuously absent from Heidegger’s account; for instance, Heidegger makes almost no mention of the body.<sup>14</sup> There is nevertheless a clear convergence of interests between Heidegger’s phenomenology and the contemporary artificial intelligence critique on at least three issues: the importance of the environment, the importance of action, and the critique of representation.

Let us start with the last of these. Descartes’ view of our relation to the world is epistemic and representational. As a result of the famous arguments he marshals at the beginning of the *Meditations*, he shows that it is impossible for us to be in direct perceptual contact with real things. The best-case scenario (after the skeptic has been dismissed) is that we have “thoughts” (in this case perceptions) that are veridical, *i.e.*, representations that in fact represent the world as it is. Whatever its argumentative merits, such an analysis presupposes that our basic *attitude* towards the world is epistemic: before anything, the question is whether we can *know* the world. And it is in particular Descartes who draws out the logical implication of this priority: that we are always essentially separated from the world by a kind of representational veil.

Heidegger does not argue that we *never* entertain an epistemic relation with the world, but he does claim that “knowing the world” is “founded” on a different and philosophically primary way of being

---

<sup>13</sup> This is Heidegger’s term of art for human beings; it designates “those beings that we ourselves are.” (SZ, 46)

<sup>14</sup> Other phenomenologists, notably Merleau-Ponty, clearly do. My story might have to be nuanced to deal with the relation of such thinkers (also regularly cited in the artificial intelligence literature). But I want to restrict my consideration to Heidegger here.

that he calls “being-in-the-world.” (SZ, 59, 61) This term is somewhat subtle, but it is not wrong to read it as suggesting that we (*Dasein*) are already in the world. Much of Division One of *Being and Time* is devoted to the presentation of phenomenological arguments designed to demonstrate that we do in fact encounter the world in the manner of prior immersion in it, and that knowing the world must be accounted for on this basis.<sup>15</sup> It follows from this that our basic way of relating to the world is not a representational one.

As for the first two convergent interests mentioned above (the importance, respectively, of the environment and of action), Heidegger’s strategy here can seem odd because he spends much of Chapter 3 elaborating a novel distinction between two categories of the being of *non-Dasein*. What he calls the “at-hand [*das Vorhandene*]” corresponds broadly to the traditional philosophical idea of an object as the correlate of a representation: a thing imbued with describable properties. But most of the time we do not encounter anything at-hand; rather we encounter things in the context of our interactions with them as equipment, that is to say as tools or tool-like things. This category of things Heidegger terms the “to-hand [*das Zuhandene*].” The most basic structure of equipment is that it has an essential “in-order-to [*Um-Zu*]” component (SZ, 68) that is ultimately cashed out in a reference to human purposes and that cannot be properly understood unless we are actively involved in using the thing (as opposed to being in the representational mode of contemplation). The to-hand is prior to the merely at-hand and is the basis for understanding the latter. But Heidegger argues at length that the phenomenology of *Dasein*’s encounters with the to-hand, while not at all the same as our “theoretical” encounters, is at the same time also not “blind.” Rather, such non-theoretical activity “has its own kind of sight” that he terms “*circumspection* [*Umsicht*]” and is therefore capable of complex tasks. (SZ, 69)

Heidegger sometimes helpfully glosses his conception of being-in-the-world as a kind of direct—*i.e.*, not representationally mediated—perceptual realism: “Proximally, what we hear is never ever noises or sound complexes, but the creaking wagon, the motorbike. We hear the column on the march, the north wind, the woodpecker knocking, the crackling fire.” Indeed, as Heidegger points out, one must adopt a “very artificial and complicated attitude” to hear “pure noise” (as the music of John Cage demonstrates), something that he takes as “phenomenal evidence” of the fact that “as being-in-the-world *Dasein*

---

<sup>15</sup> In Heidegger’s terminology (he rejects the notion of experience), *Dasein* is as being-in-the-world.



dwells amidst the innerworldly to-hand, and not...amidst 'sensations,'" *i.e.*, not amid the inner subject world of Cartesian representations. (SZ, 163–64)

On this analysis, the to-hand is intricately involved with the being of *Dasein*, and it is only on the basis of a correct phenomenological analysis of the way *Dasein* typically encounters equipment that it becomes possible to give a proper account of the being of *Dasein* as being-in-the-world. Heidegger's account thus converges with that of contemporary artificial intelligence: the environment is not just a neutral target of possible epistemic representation, but more like a set of tools that afford their own non-representational resources for (intelligent) action; and action is itself seen not as merely the final outcome of an essentially epistemic and representational process, but rather as intimately bound up with our very ability to encounter things in the world at all.

### III

In this section I will address the objection that Heidegger's deep-seated anti-naturalism makes it implausible that his critique of Descartes will be useful to the naturalist project of artificial intelligence. Heidegger's basic critique of historical understandings of the being of *Dasein* is that such understandings have always tacitly thought human being on the model of the being of non-human things, but "beings with *Dasein*'s kind of being cannot be conceived in terms of reality and substantiality" (SZ, 212), or more generally "in the sense of the *being-at-hand* of other created things." (SZ, 40) But presumably the task of any kind of naturalism in the form of a science of cognition is to do just this: to explain human beings (in their cognitive aspect, understood in the widest sense) in terms derived from the scientific study of non-*Dasein* beings.

Heidegger is indeed critical of Descartes. But the basic outline of his objection to Cartesian substance dualism is, on the face of it, far removed from the standard naturalist objections to Cartesian dualism. The lowest common factor of naturalist objections is that there is no need to posit a second (thinking) substance, *res cogitans*, in addition to extended (*i.e.*, material) substance, *res extensa*, in order to explain human cognitive functioning. Different varieties of this critique might emphasize that thinking substance is not just needless, but incoherent or otherwise implausible. And they may even follow the pattern described above, attending to the extent to which the whole Cartesian structure may need to be overhauled to properly account for the effects of doing away with thinking substance. But



the common premise of all such critiques is that the assumption of a non-naturalistic thinking substance to account for human cognitive function must be rejected.

Heidegger specifically targets the inadequacy of the Cartesian doctrine of thinking substance in his critique of Descartes' account of the being of *Dasein*. But rather than taking Descartes to task for positing the subject as something that is so radically distinct from material nature that it must be considered as a different substance, Heidegger claims that Descartes "understands [*faßt*] the being of 'Dasein'...in the same way as the being of the *res extensa*, as substance" and that *this* is what "blocks the way" to understanding how *Dasein* actually is. (SZ, 98) In other words, Heidegger believes that Descartes' understanding of the essence of human being as thinking substance is faulty because it is *too close* to an understanding of human beings as natural (extended) beings (since it shares with extended nature a determination through substance) whereas the lowest common factor of modern naturalistic critiques of Descartes takes it as basic that the mistake involves a thinking substance that is *too distant* from material nature. In a slogan, Heidegger criticizes Descartes' notion of thinking *substance* whereas naturalist critiques aim at his notion of *thinking* substance.

The anti-naturalism of Heidegger's critique of Descartes is not unusual in phenomenology more generally. The very title—*Naturalizing Phenomenology*—of an important collection of articles on phenomenology and cognitive science presupposes that phenomenology is *itself* non-naturalistic in orientation, and the editors cite Husserl on the very first page declaring that "we [phenomenologists] are fighting against the naturalization of consciousness."<sup>16</sup> The essentially transcendently idealist stance of the classical phenomenological tradition makes a *rapprochement* with the sciences of the mind very difficult. But of course various options are open that make it possible to recover the insights of phenomenology within a scientific context.

One strategy, widespread in the literature on Heidegger and cognitive science, and common to all the participants in the debate that I shall be following, is to (re)interpret the phenomenological analyses in a transcendently realist way. There is some textual evidence from *Being and Time* that Heidegger's thought is consistent with the mind-independent (or *Dasein*-independent) existence of reality. (SZ, §43) And some Heidegger commentators claim he identifies the merely at-hand with the objects of scientific descriptions of this

---

<sup>16</sup> Petitot, Varela, Pachoud and Roy, *Naturalizing Phenomenology*, 1.

mind-independent reality while strongly distinguishing it from the phenomenology of *Dasein's* everyday encounters with beings, encounters best described as of things as to-hand. (CW, 152f.) In the literature this view is controversial<sup>17</sup>, but since it is common to all those who defend Heidegger's contribution to revolutionizing artificial intelligence, I am simply going to stipulate it here. Thus: there are, *prima facie*, two layers of interest to cognitive science: an experiential layer that yields phenomenological descriptions and a causal layer that enables the experiential layer. Probably some setup like this is the most basic necessary condition for any dialogue at all between cognitive science/artificial intelligence and phenomenology. To the extent that cognitive science/artificial intelligence take themselves to be giving an explanation of something with an experiential component, phenomenology is required in the most general sense because it is the procedure by means of which the experiential component of the *explanandum* is specified.<sup>18</sup> But I take it that this view is too weak to underwrite the kind of revolution in artificial intelligence that Heidegger's phenomenology is supposed to have produced. What is needed is some way of inferring, or at least using induction, from the particular structures of the phenomenology of the to-hand to arrive at a substantive claim about the constitution of the causal layer.

In fact not all ways will do, and I want at the outset to exclude one possible way of making an inference from phenomenology to enabling structure because its view about the causal layer is too negative to induce the revolutionary scientific consequences of Heideggerianism at issue in this paper. It is based on an interpretation of Heidegger that postulates the irreducibility of to-hand characterizations to at-hand ones. This interpretation implies that we can indeed make a substantive inference from the phenomenology to causal structure: if the phenomenology is at-hand, then some causal story can at least in principle be told; but if the phenomenology is to-hand, then no causal story is accessible to us. In the literature on Heidegger and artificial intelligence the thinker that comes closest to representing this view is Hubert Dreyfus.

---

<sup>17</sup> Glazebrook, for instance, defends an allegedly scientifically realist view of Heidegger, but does so while maintaining that the at-hand is just as much a way of knowing founded in human practices as the to-hand. See Trish Glazebrook, "Heidegger and Scientific Realism," *Continental Philosophy Review*, vol. 34, no. 4 (2001), 361–401.

<sup>18</sup> Gallagher and Zahavi cite this claim as going back to Nagel, while also characterizing it, correctly in my view, as "rather trivial" (*The Phenomenological Mind*, 9, 24).

One of Dreyfus' most important original contributions is his detailed phenomenology of skill acquisition.<sup>19</sup> Dreyfus observes that it is the *acquisition* and practice of, or training in skilful behaviour that is characterized by rule following, instruction reading, deliberation, and the like (*i.e.*, the characteristics of our encounters with things as at-hand). When a skill has been acquired it is typically exercised automatically and without reflection; but (as Heidegger claims) this is not to say that the skill is somehow "mechanical" or uninteresting: chess and piano playing would both be examples of such skills. For Dreyfus the phenomenology of this kind of "skilful coping"<sup>20</sup> is primarily perceptual: one simply sees the right thing to do, and it jumps out as salient against the Background<sup>21</sup> of irrelevant material as the appropriate next step. Indeed, familiarly, skilful coping can be impeded by excessive rational deliberation, as in cases where one is worried about a difficult bit in a piece one is playing on the piano, and the very thinking about it makes it go badly.

Dreyfus regards traditional artificial intelligence as a research paradigm governed by the assumption that all aspects of human intelligent action, including the skilful coping characteristic of *Dasein's* typical encounters with the to-hand, can be exhaustively modelled using at-hand methods, *i.e.* (for Dreyfus), in an algorithmically specifiable and rule-governed way consistent with its representation as a computer program. Concomitantly, he regards the failure of traditional artificial intelligence as empirical corroboration of the soundness of Heidegger's philosophical argument that the circum-spective phenomenology characteristic of our encounters with the to-hand cannot be understood on the basis of the at-hand.

The main drift of Dreyfus' work manifests skepticism about the ability of artificial intelligence to explain instances of skilful or smooth coping. But I do not think Dreyfus has ever maintained the thesis that the phenomenology of the to-hand is irreducible to *any* causal explanation. Rather his emphasis has been on an independently grounded characterization of artificial intelligence as

---

<sup>19</sup> Hubert L. Dreyfus and Stuart E. Dreyfus, *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (New York: The Free Press, 1986), Chapter 1.

<sup>20</sup> Dreyfus, *Being-in-the-World*, 46.

<sup>21</sup> "Background" with a capital "B" is a technical term for Dreyfus, derived from the figure/(back)ground distinction, but intended to apply to an irreducibly basic feature of the perceptual fields: that it presents us saliences, "things" that stand out as affording possible actions. Everything that is not salient retreats into the "Background," but may emerge as salient at some other time. See Dreyfus, *Being-in-the-World*, Chapter 1.

exclusively interested in rule-governed processes, a characterization which leaves open—at least in theory—whether there might be a *different*, non-rule-based but still clearly causal-scientific understanding of the enabling level that, by implication, might be capable of explaining phenomenological instances of skilful coping. And indeed sometimes Dreyfus has adverted to non-traditional aspects of artificial intelligence and cognitive science, provisionally endorsing both Connectionism<sup>22</sup> and more recently some forms of Dynamical Systems Theory.<sup>23</sup>

Dreyfus makes an implicit contrast between artificial intelligence and cognitive science. He thinks of computers as essentially rule-following entities. So the claim that the phenomenology of the to-hand cannot be accounted for on the basis of rules yields the conclusion that there are *a priori* limits to what *computers* can do. But causal models are not limited to (explicit) rule-following mechanisms, so there is conceptual space for a new kind of cognitive science. It is not clear what stops this new cognitive science becoming also a new artificial intelligence, but this is not something developed by Dreyfus himself. Nevertheless, those who do postulate a positive relation between the Heideggerian phenomenological layer and the causal layer clearly build on Dreyfus' interpretation.

This can be seen in particular by considering the implications of the possible space Dreyfus opens up. If it is true that explicitly representational and consciously rule-governed tasks require traditional artificial intelligence techniques at the implementation level, then perhaps cognitive tasks whose phenomenology is circumspective require embedded, embodied techniques at the implementation level. That is, perhaps it is possible to infer from the phenomenology of a cognitive task to the structure of its causal implementation.

---

<sup>22</sup> See Dreyfus and Dreyfus, *Mind over Machine*; Hubert L. Dreyfus and Stuart E. Dreyfus, "How to Stop Worrying about the Frame Problem, even though It's Computationally Insoluble," in *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, (ed.) Z. Pylyshyn (New York: Ablex, 1987), 95–111; Hubert L. Dreyfus and Stuart E. Dreyfus, "Making a Mind versus Modelling the Brain: Artificial Intelligence at a Branch Point," in *The Philosophy of Artificial Intelligence*, (ed.) M. Boden (Oxford: Oxford University Press, 1988/1990), 309–33.

<sup>23</sup> See Hubert L. Dreyfus, "Why Heideggerian AI Failed and How Fixing it Would Require Making it More Heideggerian," *Philosophical Psychology*, vol. 20 (2007), 247–68.

#### IV

This section presents two arguments—from Beth Preston and Michael Wheeler—purporting to show the positive import of Heidegger’s critique of Descartes for the critique of the Cartesian assumptions of traditional artificial intelligence undertaken by its contemporary embodied, embedded, and enactive counterpart.

Deploying Dreyfus’ phenomenological account of skilful practical engagement, Preston infers, with Dreyfus, that in the majority of our real-time on-line encounters with things, we engage without the mediation of reflection and representation. However, where Dreyfus sometimes seems to be using this argument effectively transcendently, to block the possibility of artificial intelligence, Preston is explicit about her use of it to constructively re-engage artificial intelligence.

For instance, Dreyfus makes use of a regress argument to show that rule-governed systems at the causal level cannot account for the phenomenology associated with the emergence of perceptual saliences. Perceptual saliences (he claims) cannot be exhaustively accounted for on the basis of explicit rules and representations because the application of a rule cannot itself be (ultimately) rule-governed. Suppose it were. Then the second rule used to apply the first rule must itself be applied. This threatens an infinite regress that can only be stopped by something that is not rule-governed. Perceptual saliences are the phenomenological registration of this something.<sup>24</sup>

But where Dreyfus detects an infinite regress in the attempt to supply an explicit and rule-governed representation of the tacit Background, Preston sees an essentially empirical argument based on the computational intractability of employing explicit rules and representations. As an example she cites Marr’s theory of vision. Marr’s theory depends in part on a proof due to Shimon Ullman that three views of four non-coplanar points on a physical body provide enough information unambiguously to recover the three-dimensional structure of that body, provided that the body is rigid. If the assumption of rigidity is not made, the problem of recovering three-dimensional structure from two-dimensional patterns of ocular irradiation is much more computationally intensive and may well be intractable. The visual system works, in part, because the assumption of rigidity is largely correct. But this assumption is not

---

<sup>24</sup> See Dreyfus and Dreyfus, *Mind over Machine*, 80; and Dreyfus, *Being-in-the-World*, 18–19.

explicitly represented by the visual system. It is—in the language of embodied, embedded, and enactive artificial intelligence—distributed through a system that includes both the visual subject and the environment. The trajectory of Preston’s rejection of representation runs parallel to Dreyfus’. But whereas Dreyfus’ argument is apparently *a priori* in nature, Preston’s is based in an empirical analysis of the relative increase in computation tractability of a system that dispenses with specific rules and representations.

Drawing again on Dreyfus, Wheeler treats the to-hand and the at-hand as the phenomenological registrations of two extremes of a continuum of possible cognitive behaviours: we (humans) experience the world in the mode of pure to-handness just when (on Dreyfus’ model of skilful practical engagement) we have managed to equip ourselves with smoothly operating sensorimotor couplings that mandate no conscious attention and therefore none of its attendant manipulation of symbolic representations. This is why, as a human with the capacity for conscious attention, I can lose track of where I am when I drive home from work or choose to think about what’s for tea when I play a well-practised piece of music on the piano. As these examples (not at all remote from Wheeler’s) suggest, behaviour in the mode of pure to-handness need not be trivial. At the other end of the continuum, behaviour in the mode of pure at-handness involves explicit ratiocination mediated by the conscious manipulation of symbols for things that need no longer be present in any way to perception. Probably only adult human beings would be capable of engaging this mode of behaviour, and even they would probably engage in it only quite rarely.

The originality of Wheeler’s suggestion however lies in the fact that he claims the very great majority of the repertoire of biological behaviour—for human beings too—falls somewhere between these two extremes. Relying again on Dreyfus’ interpretation of Heidegger, Wheeler extracts a third category from §16 of *Being and Time* where Heidegger gives a number of analyses of situations in which equipment malfunctions.<sup>25</sup> For Heidegger these analyses show how we can gain phenomenological access to the Being of equipment even when our everyday dealings with it are characterized precisely by its inconspicuousness: the to-hand nature of equipment becomes conspicuous just when the equipment does not work properly, is *not* to-hand. If the road I usually take is closed, I have to pay conscious attention for a while to figure out what route to take; I may have to

---

<sup>25</sup> Dreyfus’ pioneering interpretation of *Being and Time* §16 develops the category of what he calls the “un-ready-to-hand.” See Dreyfus, *Being-in-the-World*, 70f.

consult a map, etc., and in general deploy resources beyond a subtle but merely reactive sensorimotor hookup. What Wheeler wants to infer from this kind of analysis is the existence of a distinctive third kind of cognitive system with representational components that fall short of the full symbolic representations of traditional artificial intelligence but exceed the resources of a mere sensorimotor correlation. He describes this third kind of system as “minimally representational.” (CW, 219f.)

To take the example that Wheeler uses, Franceschini *et al.* report a robot whose artificial neurology is based on that of a housefly. This robot performs a standard task set (navigating to a target light source while avoiding interposed obstacles) using only relative motion information computed on the previous time slice. The authors describe it using representational terminology, as having a “snap map,” but are at pains to emphasize the differences between this kind of representation and a traditional (philosophical) “symbolic” representation: the snap map is transient, centred on the light-sensitive array (*i.e.*, not objective, but egocentric) and intimately, even “reactively,” tied to a corresponding motor map. In the terminology that Wheeler develops it is an “action-oriented representation.”<sup>26</sup>

As this example suggests, one of the characteristics of contemporary embodied, embedded, and enactive artificial intelligence is its increased sensitivity to issues of biological plausibility. This concern dovetails with Preston’s argument about computational tractability. The Cartesian picture cannot be plausibly generalized from an evolutionary perspective because it imposes computational tasks on simple organisms that transcend their *known* computational capacities. What is known about the severe limitations of hoverfly neuroanatomy, for instance, precludes the possibility that hoverflies have a general distance computation algorithm: they just don’t have enough neural hardware to compute something so complex. Hence, the task they solve cannot be correctly described in such terms.<sup>27</sup> Here one does not have to appeal to a full-scale computational intractability *tout court*, but rather one can see Preston as giving a formal (infor-

---

<sup>26</sup> N. Franceschini, J. M. Pichon, C. Blanes and J. M. Brady, “From Insect Vision to Robot Vision,” *Philosophical Transactions: Biological Sciences*, vol. 337, no. 1281 (1992), 283–294, here 289–90; Wheeler, *Reconstructing the Cognitive World*, 9.

<sup>27</sup> See Dave Cliff, “The Computational Hoverfly: A Study in Computational Neuroethology,” in *From Animals to Animats: Proceedings of the First International Conference and the Simulation of Adaptive Behaviour (SAB90)*, (ed.) J.-A. Meyer and S. W. Wilson (Cambridge: MIT Press Bradford Books, 1991), 87–96.



mation-theoretic) account of Brooks' informal insight that SHAKEY *must* be doing the wrong thing because insects can navigate better than it even though they have *less* computational power.<sup>28</sup>

It is, however, of some significance to note about this understanding of the project of a naturalized Heideggerianism in cognitive science that its mode of argument is not based fundamentally on the structure of the phenomenological level, but on other considerations drawn from information theory, evolutionary biology, and neuroanatomy.

## V

In this section I will argue that, although there are suggestive convergences between Heidegger's phenomenology of the to-hand and contemporary practice in artificial intelligence, the phenomenology is making little or no substantive contribution to that practice.<sup>29</sup> First, there is at least one significant counter-example to the claim that phenomenological structure matches causal structure: almost all varieties of contemporary linguistics exhibit a smooth-coping phenomenology, but the underlying causal structure is, as far as the best science shows, computational. Second, the kind of causal mechanisms implied by a naturalization of Heidegger's conception of equipmentality are both implausible and constrictive. Third, there are independent non-phenomenological reasons for the contemporary critique of traditional artificial intelligence, based in information theory, evolutionary biology, and neuroanatomy that are motivating both scientific practice *and* philosophical reflection (even of an explicitly Heideggerian kind). As a result, I conclude, the Heideggerian phenomenology is doing no substantive work in artificial intelligence. I will briefly highlight each of these points in turn.

First, it is clearly not true that one can simply infer from a phenomenology of smooth coping to an enabling mechanism that also lacks rules and explicit representations. As a counter-example, one might put forward one of the most successful domains of the cogni-

---

<sup>28</sup> See Brooks, *Flesh and Machines*, 39ff., 88.

<sup>29</sup> Here I will not specifically address Wheeler's claim that most intelligent behaviour is phenomenologically un-to-hand and therefore uses his "minimal representations." It is harder to be clear about what a mapping between levels would mean in this case since both sides are relatively poorly defined in comparison with the clear cases. In any event, if the relation fails to go through for the clear cases, it will be correspondingly harder to make it work for the less clear ones.

tive revolution, linguistics. There the standard view is that the phenomenology of linguistic competence is that of smooth coping, but its implementation may indeed be in terms of rules and representations.<sup>30</sup> Even the basic data that, for instance, Chomsky brings to bear to distinguish the topic of grammar in the science of linguistics from the traditional understanding of the grammarian's task show this.<sup>31</sup> He uses the example of question-formation rules. Most of the data regarding children's experience show that English questions can be formed from assertoric statements by the simple rule: "Find the first verb and put it at the front." This rule would enable the child to map from "the mouse is running" to "is the mouse running?" But the rule breaks down for statements with embedded clauses, *e.g.*, "the man who is tall is in the room" (which would generate the ungrammatical \*"is the man who tall is in the room?"). Yet apparently children never make this mistaken overgeneralization and unhesitatingly produce correct questions (if they can understand the original sentence) according to a much more subtle rule. This kind of example is intended to defend the poverty of the stimulus argument by showing that the linguistic data children are exposed to in language learning severely underdetermines the subtlety of actual judgments of grammaticality. Here though what is interesting about these arguments is that native speakers have no particular insight into the nature and structure of the rules that determine grammaticality; indeed, the science of linguistics takes a lot of pains to recover those rules. But this shows that linguistic competence has the form of a smooth coping in which we do not formulate explicit rules and representations, but simply use language.<sup>32</sup> And yet, although the theories about its precise constitution have changed considerably over the past sixty years or so, the structures uncovered by linguistics take the form of computations performed over representations. In other words, contemporary linguistics shows that the best causal explanation for the implementation of a cognitive domain characterized phenomenologically as to-hand is precisely in terms of the rules and representations that characterize the at-hand or traditional artificial intelligence. There is almost no discussion of contemporary linguistics in any of the works under consideration, and this marks a problematic refusal to consider the important challenge it presents

---

<sup>30</sup> See Noam Chomsky, *Rules and Representations* (New York: Columbia University Press, 1980).

<sup>31</sup> Noam Chomsky, *Reflections on Language* (London: Fontana, 1975), 30f.

<sup>32</sup> Of course a phenomenological argument could be used to show the same thing.

to the view that it is possible to infer causal from phenomenological structure.

Second, the thesis that there is a strong relation between causal and phenomenological layers also has disturbing consequences in the increasingly biological context in which the science of the enabling layer is—correctly in my view—now embedded. The paradigm of the phenomenology associated with the to-hand is of course tool use. On Dreyfus' reading, a tool is a physical object belonging to a class that can only be identified on the basis of the role that the object plays in a purposive human practice. In Heidegger's canonical example, a hammer is a hammer only on the basis of our understanding of the role it plays in the practice of hammering, what he terms its *Wozu*, or "towards-which." Practices are functionally related to each other, so the nail refers to a practice whose *Wozu* is, in part, to provide an input (what Heidegger terms the *Woraus*, the out-of-which, of hammering; conversely, sheltering is something that may in part be comprised by hammering). This is the basis of Heidegger's argument that there is "strictly speaking no such thing as a piece of equipment," *i.e.*, that "equipmentality" is systematic. (SZ, 68–69)

But the nested sequences of social practices do not regress infinitely in either direction. In one direction such sequences are terminated by a direct reference to human interests (*e.g.*, sheltering) that provide the ultimate significance of equipment, its *Worumwillen*, its for-the-sake-of-which; and in the other, natural beings can provide the *Woraus* for a practice but are not themselves the result of any further practice.<sup>33</sup> (SZ, 87)

The crucial move that Wheeler makes explicit and which is required to articulate Heidegger productively with the kind of (*nouvelle*) cognitive science he promotes is to generalize this non-paradigmatic mode of the to-hand so that it can apply to non-tool-using animals and animals that lack anything that Dreyfus might identify as social practices. Wheeler therefore posits a "biological background" analogous to Dreyfus' social background: "a set of evolutionarily determined behavioural norms that constitute an externally constituted holistic network of significance within which animal behaviour is cognitively situated." Wheeler describes this as an "extension" of Dreyfus' Heidegger and also tables the possibility that animals *with* social practices (typically human beings) will re-

---

<sup>33</sup> Natural beings that do not even feature as the inputs of an equipmental practice may also be viewed as equipment, but not in a paradigmatic way. (SZ, 70) See Robert Brandom, "Heidegger's Categories in *Being and Time*," *The Monist*, vol. 66, no. 3 (1983), 387–409.

articulate what one might name their biological practices into their social practices. (CW, 159–60)

But the conceptual “extension” is warranted only on the basis of a projection of the fundamentally tool-like, *i.e.*, instrumental, structure (*Umzu*) of the social practice interpretation of the at-hand onto biological practices. Such a picture of the biological as providing a naturalistic space for instrumentality or functionality is of course adaptationism. Wheeler exhibits considerable explicit sympathy for highly adaptationist programs in evolutionary biology like evolutionary psychology<sup>34</sup> (CW, 61, 289 n. 8) and also deploys such reasoning in his (speculative) solution to the frame problem. His argument there relies on making niche-dependent sensing the general case so that “each special-purpose adaptive coupling, as selected by evolution or learning brings a context along with it, implicitly realized in the very operating principles that define the mechanism’s successful functioning.” (CW, 277)

Wheeler’s discussions of artificially evolutionary strategies in robotics also (perhaps unwittingly) underwrite this conception of evolution because, unlike biological evolution, artificial evolution depends upon the explicit postulation of a function (the so-called fitness function) against whose performance the population of artificially evolved beings is measured. Thus in biological evolution ways of value making are developed on the basis only of differential reproductive success, whereas in the artificial situation differential reproduction is calculated on the basis of a measure of performance of—adaptation to—a *specific task*.

Adaptationism however is a highly controversial position within evolutionary biology, and has been roundly criticized in an influential paper by Gould and Lewontin.<sup>35</sup> There they argue that many other factors influence the fixation of a phenotypic trait within a population than natural selection or adaptation. For instance, they point to the overall design or *Bauplan* of an organism as an important constraint on evolution. More recent developments suggest that adaptation is actually very *unusual* at the molecular level.<sup>36</sup>

---

<sup>34</sup> Anthony P. Atkinson and Michael Wheeler, “The Grain of Domains: The Evolutionary-Psychological Case against Domain-General Cognition,” *Mind and Language*, vol. 19, no. 2 (2004), 147–76.

<sup>35</sup> Steven J. Gould and Richard C. Lewontin, “The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme,” *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 205, no. 1161 (1979), 581–98.

<sup>36</sup> Motoo Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge: Cambridge University Press, 1983).

There does not seem to be any reason to tie artificial intelligence and cognitive science research to the claim that organisms are nothing more than bundles of traits, and cognitive organs nothing more than bundles of special-purpose modules.

Where—and this is the third point—there is convergence between structures of the phenomenological and causal levels, the adoption of the relevant styles of explanation at the causal level has always had an independent empirical justification, as is shown explicitly by Preston's arguments and implicitly by Wheeler's. Where, for instance, a non-representational case can be made in visual processes (weak in the case of human vision, stronger in the case of hoverfly vision), this case is made plausible not by phenomenological considerations, but by independent information-theoretic, neurological, or evolutionary considerations.

Wheeler clearly articulates the possibility that there may not be "isomorphism" between phenomenology and the enabling level, claiming that while sufficient, isomorphism is by no means necessary for the weaker relation he describes, following McDowell, as an "intelligible interplay" between levels.<sup>37</sup> (CW, 223, 128) This weaker relation requires only that phenomenological descriptions "may be enabled by wholly unmysterious causal mechanisms." (CW, 276) But this is hardly in question: no one is suggesting "mysterious" or "non-causal" mechanisms. The worry is therefore that there is in fact no *particular* relation at all beyond what is already contained in the idea of "enabling," that the causal layer makes the phenomenology possible. (CW, 234–35) Wheeler in particular wants to allow that the absence of isomorphism can yield "provisional suspicion" (CW, 234) that the enabling level is wrongly construed. But this relation is too weak to have substance. One would, anyway, have to seek out independent non-phenomenological evidence to validate or refute the suspicion. So why not dispense with the phenomenological level altogether and rely exclusively on the other evidence?<sup>38</sup>

---

<sup>37</sup> See John McDowell, "The Content of Perceptual Experience," *Philosophical Quarterly*, vol. 44, no. 175 (1994), 190–205, here 197.

<sup>38</sup> Indeed, in the cases typically preferred by the theoreticians of Heideggerian cognitive science—Brooksian ones of non-human real-time locomotion—there is no real question of phenomenology at all since the putative phenomenological states are inaccessible.

## Conclusion

As a result I find it hard to see exactly what positive work a close Heideggerian phenomenological analysis is doing for cognitive science or artificial intelligence, beyond the minimal position that if the target of a cognitive science is (in part) the explanation of cognitive phenomena, then some phenomenology is required to specify the ultimate *explanada* of such a science. The most interesting results of “Heideggerian” artificial intelligence and cognitive science have emerged without much philosophical consideration at all. Where philosophers have reflected on these results, it seems to me that, as with Preston, the fundamental structure of their argumentation relies not on *a priori* phenomenological premises, but on very broad information-theoretic considerations. So not only is there no strong relation (like isomorphism) between phenomenological and causal levels, but there appear always to be independent non-phenomenological reasons for postulating ‘embodied, embedded’ causal explanations where they turn out to be appropriate. As a result it is hard to see that there is any substantive contribution Heideggerian phenomenology can make to understanding causal structure beyond careful description of the *explanandum*, and tentative and empirically defeasible hypothesis suggestion.

*Alistair.Welchman@utsa.edu*