On the Role of AI in the Ongoing Paradigm Shift within the Cognitive Sciences

Tom Froese

CSRP 592

June 2007

ISSN 1350-3162

Cognitive Science
Research Papers

On the role of AI in the ongoing paradigm shift within the cognitive sciences

Tom Froese

Centre for Computational Neuroscience and Robotics (CCNR)
Centre for Research in Cognitive Science (COGS)
University of Sussex, Brighton BN1 9QH, UK

t.froese@sussex.ac.uk

Abstract

This paper supports the view that the ongoing shift from orthodox to embodied-embedded cognitive science has been significantly influenced by the experimental results generated by AI research. Recently, there has also been a noticeable shift toward enactivism, a paradigm which radicalizes the embodied-embedded approach by placing autonomous agency and lived subjectivity at the heart of cognitive science. Some first steps toward a clarification of the relationship of AI to this further shift are outlined. It is concluded that the success of enactivism in establishing itself as a mainstream cognitive science research program will depend less on progress made in AI research and more on the development of a phenomenological pragmatics.

Keywords: AI, cognitive science, paradigm shift, enactivism, phenomenology.

To appear in: M. Lungarella, F. Iida, J. Bongard & R. Pfeifer (eds.), *Proc. of the 50th Anniversary Summit of Artificial Intelligence,* Berlin, Germany: Springer-Verlag

1. Introduction

Over the last two decades the field of artificial intelligence (AI) has undergone some significant developments (Anderson 2003). Good old-fashioned AI (GOFAI) has faced considerable problems whenever it attempts to extend its domain beyond simplified "toy worlds" in order to address context-sensitive real-world problems in a robust and flexible manner[1]. These difficulties motivated the Brooksian revolution toward an embodied and situated robotics in the early 1990s (Brooks 1991). Since then this approach has been further developed (e.g. Pfeifer & Scheier 1999; Pfeifer 1996; Brooks 1997), and has also significantly influenced the emergence of a variety of other successful methodologies, such as the dynamical approach (e.g. Beer 1995), evolutionary robotics (e.g. Harvey *et al.* 2005; Nolfi & Floreano 2000), and organismically-inspired robotics (e.g. Di Paolo 2003). These approaches are united by the claim that cognition is best understood as embodied and embedded in the sense that it emerges out of the dynamics of an extended brain-body-world systemic whole.

These developments make it evident that the traditional GOFAI mainstream, with its emphasis on perception as representation and cognition as computation, is being challenged by the establishment of an alternative paradigm in the form of embodied-embedded AI. How is this major shift in AI related to the ongoing paradigm shift[2] within the cognitive sciences? Section 2 analyzes the role of AI in the emergence of what has been called "embodied-embedded" cognitive science (e.g. Clark 1997; Wheeler 2005). Recently, there has also been a noticeable shift in interest toward "enactivism" (e.g. Thompson 2007; 2005; Di Paolo, Rohde & De Jaegher 2007; Torrance 2005), a paradigm which radicalizes the embodied-embedded approach by placing autonomous agency and lived subjectivity at the heart of cognitive science. How AI relates to this further shift is still in need of clarification, and section 3 provides some initial steps in this direction. Finally, section 4 argues that since many of the claims of enactivism are grounded in the phenomenological domain, its success as a major cognitive science research program depends less on progress in AI research and more on the development of a phenomenological pragmatics.

2. Toward embodied-embedded cognitive science

Much of contemporary cognitive science owes its existence to the founding of the field of AI in the late 1950s by the likes of Herbert Simon, Marvin Minsky, Allen Newell, and John McCarthy. These researchers, along with Noam Chomsky, put forth ideas that were to become the major guidelines for the computational approach which has dominated the cognitive sciences since its inception[3]. In order to determine the impact of AI on the ongoing shift from such orthodox computationalism toward embodied-embedded cognitive science, it is necessary to briefly consider some of the central claims associated with these competing theoretical frameworks.

---

[1] For example: the commonsense knowledge problem (Dreyfus 1991, p. 119), the frame problem (McCarthy & Hayes 1969), and the symbol grounding problem (Harnard 1990).

[2] Whether any of the major changes in AI or cognitive science are paradigm shifts in the strict Kuhnian sense is an interesting question but beyond the scope of this paper. Here the notion is used in the more general sense of a major shift in experimental practice and focus.

[3] See Boden (2006) for an extensive overview of the history of cognitive science.

2.1 Theories of cognition

The paradigm that came into existence with the birth of AI, and which was essentially identified with cognitive science itself for the ensuing three decades and which still represents the mainstream today, is known as *cognitivism* (e.g. Fodor 1975). The cognitivist claim, that cognition is a form of computation (i.e. information processing through the manipulation of symbolic representations), is famously articulated in the Physical-Symbol System Hypothesis which holds that such a system "has the necessary and sufficient means for general intelligent action" (Newell & Simon 1976). From the cognitivist perspective cognition is essentially centrally controlled, disembodied, and decontextualized reasoning and planning as epitomized by abstract problem solving. Accordingly, the mind is conceptualized as a digital computer and cognition is viewed as fundamentally distinct from the embodied action of an autonomous agent that is situated within the continuous dynamics of its environment.

The cognitivist orthodoxy remained unchallenged until *connectionism* arose in the early 1980s (e.g. McClelland, Rumelhart *et al.* 1986). The connectionist alternative views cognition as the emergence of global states in a network of simple components, and promises to address two shortcomings of cognitivism, namely by 1) increasing efficiency through parallel processing, and 2) achieving greater robustness through distributed operations. Moreover, because it makes use of artificial neural networks as a metaphor for the mind, its theories of cognition are often more biologically plausible. Nevertheless, connectionism still retains many cognitivist commitments. In particular, it maintains the idea that cognition is essentially a form of information processing in the head which converts a set of inputs into an appropriate set of outputs in order to solve a given problem. In other words, "connectionism's disagreement with cognitivism was over the nature of computation and representation (symbolic for cognitivists, subsymbolic for connectionsists)" (Thompson 2007, p. 10), rather than over computationalism as such (see also Wheeler 2005, p. 75). Accordingly, most of connectionism can be regarded as constituting a part of orthodox cognitive science.

Since the early 1990s this computationalist orthodoxy has begun to be challenged by the emergence of *embodied-embedded* cognitive science (e.g. Varela, Thompson & Rosch 1991; Clark 1997; Wheeler 2005), a paradigm which claims that an agent's embodiment is constitutive of its perceiving, knowing and doing (e.g. Gallagher 2005; Noë 2004; Thompson & Varela 2001). Furthermore, the computational hypothesis has given way to the dynamical hypothesis that cognitive agents are best understood as dynamical systems (van Gelder 1998). Thus, while the embodied-embedded paradigm has retained the connectionist focus on self-organizing dynamic systems, it further holds that cognition is a situated activity which spans a systemic totality consisting of an agent's brain, body, and world (e.g. Beer 2000). In order to assess the importance of AI for this ongoing shift toward embodied-embedded cognitive science, it is helpful to first consider the potential impact of theoretical argument alone.

2.2 A philosophical stalemate

The theoretical premises of orthodox and embodied-embedded cognitive science can generally be seen as Cartesian and Heideggerian in character, respectively (e.g. Wheeler 2005; Dreyfus 2007; Anderson 2003). The traditional Cartesian philosophy accepts the assumption that any kind of being can be reduced to a combination of

more basic atomic elements which are themselves irreducible. On this view cognition is seen as a general-purpose reasoning process by which a relevant representation of the world is assembled through the appropriate manipulation and transformation of basic mental states (Wheeler 2005, p. 38). Orthodox cognitive science adopts a similar kind of reductionism in that it assumes that symbolic/subsymbolic structures are the basic representational elements which ground all mental states[4], and that cognition is essentially treated as the appropriate computation of such representations. What are the arguments against such a position?

The Heideggerian critique starts from the phenomenological claim that the world is first and foremost experienced as a significant whole and that cognition is grounded in the skilful disposition to respond flexibly and appropriately as demanded by contextual circumstances. Dreyfus (1991, p. 117) has argued that such a position questions the validity of the Cartesian approach in two fundamental ways. First, the claim of *holism* entails that the isolation of a specific part or element of our experience as an atomic entity appears as secondary because it already presupposes a background of significance as the context from which to make the isolation. From this point of view a reductionist attempt at reconstructing a meaningful whole by combining isolated parts appears nonsensical since the required atomic elements were created by stripping away exactly that contextual significance in the first place. As Dreyfus (1991, p. 118) puts it: "Facts and rules are, by themselves, meaningless. To capture what Heidegger calls significance or involvement, they must be *assigned relevance*. But the predicates that must be added to define relevance are just more meaningless facts". From the Heideggerian perspective it therefore appears that the Cartesian position is faced with a problem of infinite regress. Second, if we accept the claim of *skills*, namely that cognition is essentially grounded in a kind of skilful know-how or context-sensitive coping, then the orthodox aim of reducing such behaviour into a formal set of input/output mappings which specify the manipulation and transformation of basic mental states appears to be hopelessly misguided.

Judging from these philosophical considerations it seems that the Heideggerian critique of the Cartesian tradition could have a significant impact on the paradigm shift from orthodox toward embodied-embedded cognitive science. However, since the two approaches have distinct underlying constitutive assumptions (e.g. reductionism vs. holism), there exists no *a priori* theoretical argument which would force someone holding a Cartesian position to accept the Heideggerian critique from holism and skills. Similarly, it is not possible for the Cartesian theorist to prove that worldly significance can indeed be created through the appropriate manipulation and transformation of abstract and de-contextualized representational elements. The problem is that, like all rational arguments, both accounts of cognition are founded on a particular set of premises which one is at liberty to accept or reject. Thus, even if the development of a strong philosophical position is most likely a necessary factor in the success of the embodied-embedded paradigm, it is by itself not sufficient. In other words, there is a fundamental stalemate in the purely *philosophical* domain; a shift in constitutive assumptions cannot be engendered by argumentation alone.

---

[4] In contrast to the Cartesian claim that mental stuff is ontologically basic, orthodox cognitive science could hold that these constitutive elements are not basic in this absolute sense because they are further reducible to physical states. However, this change in position does not make any difference with regard to Heidegger's critique of this kind of reductionism.

2.3 An empirical resolution

It has often been proposed that this theoretical stalemate has to be resolved in the empirical domain of the cognitive sciences (e.g. Dreyfus & Dreyfus 1988; Clark 1997, p. 169; Wheeler 2005, p. 187). The authors of the Physical-Symbol System Hypothesis (Newell & Simon 1976) and the Dynamical Hypothesis (van Gelder 1998) are also in agreement that only sustained empirical research can determine whether their respective hypotheses are viable. Research in AI[5] is thereby awarded the rather privileged position of being able to help resolve theoretical disputes which have plagued the Western philosophical tradition for decades if not centuries. This reciprocal relationship between AI and theory has been captured with the slogan "understanding by building" (e.g. Pfeifer 1996; Pfeifer & Scheier 1999, p. 299).

In what way has AI research managed to fulfil this role? Dreyfus (1991, p. 119), for example, has argued that the Heideggerian philosophy of cognition has been vindicated because GOFAI faces significant difficulties whenever it attempts to apply its Cartesian principles to real-world situations which require robust, flexible, and context-sensitive behavior. In addition, he demonstrates that the Heideggerian arguments from holism and skills can provide powerful explanations of why this kind of AI has to wrestle with the frame and commonsense knowledge problems. In a similar vein, Wheeler (2005, p. 188) argues compellingly that the growing success of embodied-embedded AI provides important experimental support for the shift toward a Heideggerian position in cognitive science. He argues that Heidegger's claim that a cognitive agent is best understood from the perspective of "being-in-the-world" is put to the test by embodied-embedded AI experiments which investigate cognition as a dynamical process which emerges out of a brain-body-world systemic whole.

2.4 The failure of embodied-embedded AI?

In light of these developments it seems fair to say that AI can have a significant impact on the ongoing shift from orthodox toward embodied-embedded cognitive science. However, while embodied-embedded AI has managed to overcome some of the significant challenges faced by traditional GOFAI, it has also started to encounter some of its own limitations. Considering the seemingly insurmountable challenge to make the artificial agents of current embodied-embedded AI behave in a more robust, flexible, and generally more life-like manner, particularly in the way that more complex living organisms do, Brooks (1997) was led to entertain the following sceptical reflections: "Perhaps we have all missed some organizing principle of biological systems, or some general truth about them. Perhaps there is a way of looking at biological systems which will illuminate an inherent necessity in some aspect of the interactions of their parts that is completely missing from our artificial systems. […] I am suggesting that perhaps at this point we simply do not *get it*, and that there is some fundamental change necessary in our thinking" (Brooks 1997). Has the field of AI managed to find the missing "juice" of life in the past decade?

---

[5] It is worth noting that there are compelling arguments for claiming that the results generated by AI research are not "empirical" in the same way as those of the natural sciences, and that this is likely to weaken their impact outside the field. Nevertheless, it is still the case that AI can provide "valuable tools for re-organising and probing the internal consistency of a theoretical position" (Di Paolo, Noble & Bullock 2000).

The existential philosopher Dreyfus, while mostly known in the field for his scathing criticisms of GOFAI, has recently referred to the current work in embodied-embedded AI as a "failure" (Dreyfus 2007). He points to the lack of "a model of our particular way of being embedded and embodied such that what we experience is significant for us in the particular way that it is. That is, we would have to include in our program a model of a body very much like ours". Similarly, Di Paolo (2003) argues that embodied-embedded robots, while in many respects an improvement over traditional GOFAI, "can never be truly autonomous. In other words the presence of a closed sensorimotor loop *does not* fully solve the problem of meaning in AI". These problems are even further amplified because, while embodied-embedded AI has focused on establishing itself as a vialbe alternative to the traditional computational paradigm, relatively little effort has been made to connect its experimental work with theories outside the field of AI, such as with theoretical biology, in order to address issues of autonomy and embodiment (Ziemke 2007). It thus seems that slowly there is an awareness growing in the field of embodied-embedded AI that something crucial is still missing in the current implementations of autonomous systems, and that this shortcoming is likely related to their particular manner of embodiment. But what could this elusive factor be? In order to answer this question we need to shift our focus back to recent developments in the cognitive sciences.

3. Further: the shift toward enactivism

The enactive paradigm originally emerged as a part of embodied-embedded cognitive science in the early 1990s with the publication of the influential book *The Embodied Mind* (Varela, Thompson & Rosch 1991). It has recently distinguished itself by more explicitly placing autonomous agency in addition to lived subjectivity at the heart of cognitive science (e.g. Thompson 2007; Thompson 2005; Di Paolo, Rohde & De Jaegher 2007). How AI relates to this further shift in the cognitive sciences is still in need of clarification. This section provides some initial steps in this direction by considering how AI can contribute to the enactive account of how our bodily activity relates to the subjective mind at three interrelated "dimensions of embodiment": 1) bodily self-regulation, 2) sensorimotor coupling, and 3) intersubjective interaction (Thompson & Varela 2001). While the development of such fully enactive AI is a significant challenge to existing methodologies, it has the potential of providing a fresh perspective on some of the issues currently faced by embodied-embedded AI.

3.1 Bodily self-regulation

This dimension of embodiment is central to the enactive account of autonomy. Since embodied-embedded AI has always been involved in extensive studies of autonomous systems (e.g. Pfeifer & Scheier 1999), it might seem that such research is particularly destined to relate to enactivism in a mutually informative manner. Unfortunately, things are not as straightforward; the enactive paradigm has a very different view of what constitutes autonomy when compared to most embodied-embedded AI (Froese, Virgo & Izquierdo 2007). Its approach can be traced to the notion of *autopoiesis*, a systems concept which originated in the theoretical biology of the 1970s (e.g. Maturana & Varela 1980). Enactivism defines an autonomous agent as a precarious self-producing network of processes which constitutes its own identity; the paradigmatic example being a living organism. Drawing from the bio-philosophy of

Hans Jonas (1966), it is claimed that such an autonomous system, one whose being is its own doing, should be conceived of as an individual in its own right, and that this process of self-constitution brings forth, in the same stroke, what is other, namely its world (e.g. Thompson 2007, p. 153). In other words, it is proposed that the continuous reciprocal process, which constitutes the autonomous system as a distinguishable individual, also furnishes it with an intrinsically meaningful perspective on its environment, i.e. autonomy lies at the basis of *sense-making* (Weber & Varela 2002).

It follows from these considerations that today's AI systems are not autonomous in the enactive sense. They do not constitute their own identity, and the only "identity" which they can be said to possess is projected onto them by the observing researcher (Ziemke 2007). The popular methodology of evolutionary robotics, for example, presupposes that an "individual" is already defined by the experimenter as the basis for selection by the evolutionary algorithm, and in the dynamical approach to AI it is up to the investigator to distinguish which subpart of the systemic whole actually constitutes the "agent" (Beer 1995). The enactive notion of autonomous agency therefore poses a significant difficulty for current AI methodologies. Nevertheless, it is worth noting that AI researchers do not have to synthesize actual living beings in order for their work to provide some relevant insights into the dimension of bodily self-regulation. Following Di Paolo (2003), a first step would be to investigate artificial systems with some self-sustaining dynamic structures. In this manner embodied-embedded AI can move beyond its current focus on closed sensorimotor feedback loops by implementing systems which have a reciprocal link between internal organization and external behaviour. Indeed, there are signs that a shift toward more concern with bodily self-regulation is starting to develop. This is demonstrated by an increasing interest in homeostasis as a regulatory mechanism for investigating, for example, sensory inversion (e.g. Di Paolo 2003), the emergence of sensorimotor coupling (e.g. Ikegami & Suzuki forthcoming), behavioural preference (e.g. Iizuka & Di Paolo forthcoming), and active perception (e.g. Harvey 2004).

3.2 Sensorimotor coupling and intersubjective interaction

Since sensorimotor embodiment is the research target of most current embodied-embedded AI, its results can have an impact on this aspect of enactivism. However, since the vast majority of such work is not concerned with how the constraints of constitutive autonomy are related to the emergence of sensorimotor behavior, it is not contributing to the enactive account of how an autonomous agent is able to bring forth its own cognitive domain. To become more relevant in this respect, the field needs to adapt its methodologies so as to deal with the enactive proposal that an agent's sense-making is grounded in the active regulation of ongoing sensorimotor coupling in relation to the viability of a precarious, dynamically self-sustaining identity (Weber & Varela 2002). This is an area which has been practically unexplored, although some promising work has begun (e.g. Ikegami & Suzuki forthcoming; Di Paolo 2003).

These considerations can be extended to the domain of intersubjective interaction, since this dimension of embodiment also involves distinctive forms of sensorimotor coupling (Thompson & Varela 2001). An enactive account of social understanding based on this continuity has recently been outlined by Di Paolo, Rohde and De Jaegher (2007). They make the important suggestion that the traditional focus on the embodiment of individual interactors needs to be complemented by an investigation

of the interaction process that takes place between them. This shift in focus enables them to extend the enactive notion of sense-making into the realm of social cognition in the form of *participatory sense-making*. The development of such an account is important for embodied-embedded AI, because most of its current research remains limited to "lower-level" cognition. Exploring the domain of social interaction might provide it with the necessary means to tackle the problem of scalability (Clark 1997, p. 101), in particular because such inter-action can constitute new ways of sense-making that are not available to the individual alone. The challenge is to implement AI systems that constitute the social domain by means of an interaction process that is essentially embodied and situated, as opposed to the traditional means of formalized transmissions of abstract information over pre-specified communication channels. Di Paolo, Rohde and De Jaegher (2007) review some initial work in this direction which demonstrates that "these models have the possibility to capture the rich dynamics of reciprocity that are left outside of traditional individualistic approaches".

3.3 A fully enactive AI?

It is debatable if AI research should be considered as enactive rather than embodied-embedded if it does not address some form of bodily self-regulation[6]. In this sense the authors of *The Embodied Mind* perhaps got slightly carried away when they referred to the emergence of Brooks's behaviour-based robotics as a "fully enactive approach to AI" (Varela, Thompson & Rosch 1991, p. 212). However, this is not to say that embodied-embedded AI does not have an impact on the shift toward enactivism, it does, but only to the extent that there is an overlap between the two paradigms. Its current influence is therefore by no means as significant as it has been on the shift toward embodied-embedded cognitive science. For example, Thompson's recent book *Mind in Life*, which can be considered as a successor to *The Embodied Mind*, does not even include AI as one of the cognitive science disciplines from which it draws its insights (Thompson 2007, p. 24). Indeed, at the moment it seems more likely that the influence will run more strongly from enactive cognitive science to AI instead. Its account of *autonomous agency*, for example, has the potential to provide embodied-embedded AI with exactly the kind of bodily organizational principle that has been identified as missing by Brooks (1997). In addition, the enactive notion of *sense-making*, as a biologically grounded account of how a system must be embodied in order for its encounters to be experienced as significant, can be used as a response to Dreyfus's (2007) vague requirement of "a detailed description of our body", which apparently has not even "a chance of being realized in the real world". Furthermore, there is a good possibility that the field's current restriction to "lower-level" cognition could be overcome in a principled manner by extending its existing research focus on sensorimotor embodiment to also include *participatory sense-making*.

Of course, it goes without saying that all of these aspects of enactivism are also open to further refinement through artificial modelling, and that some initial work in this direction has already begun. Nevertheless, for AI to have a more significant impact on the ongoing shift toward enactive cognitive science, it must address some considerable challenges that face its current methodologies. The field needs to extend

---

[6] In a similar manner it could be argued that since recent work in enactive perception (e.g. Noë 2004) is more concerned with sensorimotor contingencies than with autonomous agency or lived subjectivity, such work might be more usefully classified as part of embodied-embedded cognitive science rather than enactivism proper. See also Thompson (2005).

its current preoccupation with sensorimotor interaction in the behavioural domain to include a concern of the constitutive processes that give rise to that domain in living systems. Maybe Brooks (1997) was right when he suggested that in order for AI to be more life-like perhaps "there is some fundamental change necessary in our thinking". At least such a change is indeed necessary for the development of a fully enactive AI.

4. From AI to phenomenology

How can such fully enactive AI impact on the cognitive sciences? This section argues that, while clearly an important aspect, it is not sufficient to displace the orthodox mainstream. More than just having to make Heideggerian AI more Heideggerian, as Dreyfus (2007) proposes, Heideggerian cognitive science itself must become more Heideggerian by shifting its focus from AI to phenomenology, a shift which coincides with a movement from embodied-embedded cognitive science to enactivism.

4.1 An empirical stalemate

Two decades ago Dreyfus and Dreyfus (1988) characterized GOFAI as a project in which "the rationalist tradition had finally been put to an empirical test, and it had failed". Nevertheless, despite this supposed 'failure' no alternative has yet succeeded in fully displacing the orthodox mainstream in AI or cognitive science. While it could be argued that more progress in embodied-embedded or enactive AI will eventually remedy this situation, a more serious problem becomes apparent when we consider why this perceived 'failure' did not remove the orthodox framework from the mainstream. As Wheeler (2005, p. 185) points out, this did not happen for the simple reason that researchers are always at liberty to interpret practical problems as mere temporary difficulties which will eventually be eliminated through more scientific research and additional technological development. Accordingly, Wheeler goes on to conclude that a resolution of the standoff must await further empirical evidence.

However, while Wheeler's appeal to more experimental data is evidently useful when resolving theoretical issues within a particular approach, it is not clear whether it is also valid when deciding between different paradigms: you always already have to choose (whether explicitly or not) one paradigm over the others from which to interpret the data. Furthermore, this choice is significant because "the conceptual framework that we bring to the study of cognition can have profound empirical consequences on the practice of cognitive science. It influences the phenomena we choose to study, the questions we ask about these phenomena, the experiments we perform, and the ways in which we interpret the results of these experiments" (Beer 2000). Thus, since data is only meaningful in a manner which crucially depends on the underlying premises of the investigator, the current empirical stalemate in AI appears to be less due to a lack of empirical evidence and more due to the fact that the impact of an experiment fundamentally depends on an interpretive aspect[7]. In other words, in order for experimental *data* to be turned into scientific *knowledge* it first has to be *interpreted* according to (often implicitly) chosen constitutive assumptions. Moreover, our premises even ground the manner in which we distinguish between

[7] Again, this is not to say that such experimental evidence has no effect. The point is simply that, while a necessary component, it is not *sufficient* for a successful paradigm shift.

noise and data[8]. It follows from this that the major cause of the standoff in the philosophical domain also plays a significant role in the current empirical stalemate: *both domains of enquiry require an interpretative action on the part of the observer*. And, more importantly, while it is possible to influence this act of interpretation through research progress, its outcome cannot be fully determined by such external events since any kind of understanding always already presupposes interpretative activity[9]. In addition, the impact of this potential influence is also limited because the significance of such advances might not become apparent if one does not already hold the kind of constitutive assumptions required to understand them appropriately.

These considerations give a rather bleak outlook for the possibility of actively generating a successful paradigm shift in the cognitive sciences, and at this point it might seem relatively futile to worry about such abstract problems and just get on with the work. Indeed, considering the overall state of affairs this is in many respects a sensible and pragmatic course of action. Nevertheless, it is evidently the case that we choose a paradigm for our research. However, if rational argument combined with empirical data is still not sufficient to necessarily establish this choice, then what is it that determines which premises are assumed? And how can this elusive factor be influenced? The rest of this section provides a tentative answer to these questions by focusing on a crucial aspect of enactivism that has not been addressed so far.

4.2 A phenomenological resolution

The enactive account of autonomous agency as expressed in terms of systems biology is complemented by a concern with the first-person point of view, by which is meant the subjectively *lived experience* associated with cognitive and mental events (Varela & Shear 1999). Since the enactive framework incorporates both biological agency and phenomenological subjectivity, it allows the traditional mind-body problem to be recast in terms of what has recently been called the "body-body problem" (Hanna & Thompson 2003). On this view the traditional "explanatory gap" (Levine 1983) is no longer absolute since the concepts of subjectively lived body and objective living body both require the notion of living being. Though more work needs to be done to fully articulate the details, this reformulation of the hard problem of consciousness can be seen as one of the major contributions of enactivism (Torrance 2005).

Nevertheless, it is not yet clear how a concern with subjective experience could provide us with a way to move beyond the stalemate that we have identified in the previous sections. Surely enactivism is just more philosophical theory? However, to say this is to miss the point that it derives many of its crucial insights from a source that is quite distinct from standard theoretical or empirical enquiry, namely from careful *phenomenological* observations that have been gained through the principled investigation of the structure of our lived experience (see Ch. 2 in Thompson 2007 for an overview). But what about the insights from which Heidegger originally deduced

---

[8] Consider, for example, the empirical fact that the fossil record shows long periods of stasis interspersed with layers of rapid phyletic change. Someone who believes that evolution proceeds gradually will treat this fact as irrelevant noise, while someone who claims that evolution proceeds as punctuated equilibria will view such a finding as supporting evidence.

[9] From the enactive view this is hardly surprising (Varela, Thompson & Rosch 1991, p. 10-12), and it can ground this epistemological reflection in its biology of autonomy by claiming that a living system always constitutes its own perspective on the world. Indeed, at one point enactivism was actually called "the hermeneutic approach" (Thompson 2007, p. 24).

his claims? If his analysis of the holistic structure of our "being-in-the-world" is one of the most influential accounts of the Husserlian phenomenological tradition, then why did it not succeed in convincing mainstream cognitive scientists? The regrettable answer is that, while his claims have sometimes been probed in the philosophical or empirical domain, there have not been many sustained and principled efforts in orthodox cognitive science to verify their validity in the phenomenological domain.

If enactivism is to avoid this fate then it needs to focus less on the development of enactive AI, and more on the promotion of principled phenomenological studies. Indeed, according to Di Paolo, Rohde and De Jaegher (2007) the central importance of experience is perhaps the most revolutionary implication of enactivism since "phenomenologically informed science goes beyond black marks on paper or experimental procedures for measuring data, and dives straight into the realm of personal experience" such that, for example, "no amount of rational argument will convince a reader of Jonas's claim that, as an embodied organism, he is concerned with his own existence if the reader cannot see this for himself". Thus, enactivism implicates an element of personal practice. Similarly, Varela and Shear (1999) outline the beginnings of a project "where neither experience nor external mechanism have the final word", but rather stand to each other in a relationship of mutual constraints. They point out that the collection of phenomenological data requires a disciplined training in the skilful exploration of lived experience. Such an endeavour might already be worthwhile in itself, but in the context of the stalemate in the cognitive sciences it comes with an added benefit. In a nutshell this is because, while it is still the case that phenomenological data first has to be interpreted from a particular point of view before it can be integrated into a conceptual framework, generating such data also requires a change in our mode of experiencing. Moreover, this change in our experiential attitude is constituted by a change in our mode of *being*, and this in turn entails a change in our *understanding* (Varela 1976). Thus, it is this being, our everyday "Dasein", which determines how we interpret our world. Of course, since we are autonomous agents this does not mean that actively practicing phenomenology necessarily commits us to enactivism. But perhaps by changing our awareness in this manner we will be able to understand more fully the reasons, other than theory and empirical data, which are at the root of why we prefer one paradigm over another.

5. Conclusion

The field of AI has had a significant impact on the ongoing shift from orthodox toward embodied-embedded cognitive science mainly because it has made it possible for philosophical disputes to be addressed in an experimental manner. Conversely, enactivism can have a strong influence on AI because of its biologically grounded account of autonomous agency and sense-making. The development of such enactive AI, while challenging to current methodologies, has the potential to address some of the problems currently in the way of significant progress in embodied-embedded AI. However, if an alternative paradigm is to be successful in actually displacing the orthodox mainstream, then it is unlikely that theoretical arguments and empirical evidence alone are sufficient. For this to happen it will be necessary that a *phenomenological pragmatics* is established as part of the general methodological toolbox of contemporary cognitive science. This shift of focus from AI to phenomenology coincides with a shift from embodied-embedded cognitive science to

enactivism. Unfortunately, however, most of our current academic institutions are not concerned with supporting phenomenology in any principled manner, and it will be one of the major challenges facing the realization of enactivism as the mainstream of cognitive science to devise appropriate ways of changing this. In this context, Terry Winograd's turn toward teaching Heidegger in computer science courses at Stanford when he became disillusioned with traditional GOFAI appears in a new light[10].

References

Anderson, M.L. (2003), "Embodied Cognition: A field guide", *Artificial Intelligence*, **149**(1), pp. 91-130

Beer, R.D. (1995), "A dynamical systems perspective on agent-environment interaction", *Artificial Intelligence*, **72**(1-2), pp. 173-215

Beer, R.D. (2000), "Dynamical approaches to cognitive science", *Trends in Cognitive Sciences*, **4**(3), pp. 91-99

Boden, M.A. (2006), *Mind as Machine: A History of Cognitive Science*, 2 vols., Oxford, UK: Oxford University Press

Brooks, R.A. (1991), "Intelligence without representation", *Artificial Intelligence*, **47**(1-3), pp. 139-160

Brooks, R.A. (1997), "From earwigs to humans", *Robotics and Autonomous Systems*, **20**(2-4), pp. 291-304

Clark, A. (1997), *Being There*, Cambridge, MA: The MIT Press

Di Paolo, E.A. (2003), "Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop", in: K. Murase & T. Asakura (eds.), *Dynamical Systems Approach to Embodiment and Sociality*, Adelaide, Australia: Advanced Knowledge International, pp. 19-42

Di Paolo, E.A., Noble, J. & Bullock, S. (2000), "Simulation Models as Opaque Thought Experiments", in: M.A. Bedau *et al.* (eds.), *Proc. of the 7th Int. Conf. on the Synthesis and Simulation of Living Systems*, Cambridge, MA: The MIT Press, pp. 497-506

Di Paolo, E.A., Rohde, M. & De Jaegher, H. (2007), "Horizons for the Enactive Mind: Values, Social Interaction, and Play", Cognitive Science Research Paper, **587**, University of Sussex, Brighton, UK

Dreyfus, H.L. (1991), *Being-in-the-World*, Cambridge, MA: The MIT Press

Dreyfus, H.L. (2007), "Why Heideggerian AI failed and how fixing it would require making it more Heideggerian", *Philosophical Psychology*, **20**(2), pp. 247-268

[10] See Dreyfus (1991, p. 119).

Dreyfus, H.L. & Dreyfus, S.E. (1988), "Making a mind versus modelling the brain: artificial intelligence back at a branch-point", *Daedalus*¸ **117**(1), p. 15-44

Fodor, J.A. (1975), *The Language of Thought*, Cambridge, MA: Harvard Uni. Press

Froese, T., Virgo, N. & Izquierdo, E. (2007), "Autonomy: a review and a reappraisal", in: F. Almeida e Costa *et al.* (eds.), *Proc. of the 9th Euro. Conf. on Artificial Life*, Berlin, Germany: Springer-Verlag, in press

Gallagher, S. (2005), *How the Body Shapes the Mind*, New York, NY: Oxford University Press

Hanna, R. & Thompson, E. (2003), "The Mind-Body-Body Problem", *Theoria et Historia Scientarum*, **7**(1), pp. 24-44

Harnard, S. (1990), "The symbol grounding problem", *Physica D*, **42**, pp. 335-346

Harvey, I. (2004), "Homeostasis and Rein Control: From Daisyworld to Active Perception", in: J. Pollack *et al.* (eds.), *Proc. of the 9th Int. Conf. on the Simulation and Synthesis of Living Systems*, Cambridge, MA: The MIT Press, pp. 309-314

Harvey, I., Di Paolo, E.A., Wood, R., Quinn, M. & Tuci, E. A. (2005), 'Evolutionary Robotics: A new scientific tool for studying cognition', *Artificial Life*, **11**(1-2), pp. 79-98

Iizuka, H. & Di Paolo, E.A. (forthcoming), "Toward Spinozist robotics: Exploring the minimal dynamics of behavioral preference", *Adaptive Behavior*

Ikegami, T. & Suzuki, K. (forthcoming), "From Homeostatic to Homeodynamic Self", *BioSystems*

Jonas, H. (1966), *The Phenomenon of Life: Toward a Philosophical Biology*, Evanston, Illinois: Northwestern University Press, 2001

Levine, J. (1983), "Materialism and Qualia: The Explanatory Gap", *Pacific Philosophical Quarterly*, **64**, pp. 354-361

Maturana, H.R. & Varela, F.J. (1980), *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht, Holland: Kluwer Academic Publishers

McCarthy, J. & Hayes, P.J. (1969), "Some philosophical problems from the standpoint of artificial intelligence", in: B. Meltzer & D. Michie (eds.), *Machine Intelligence 4*, Edinburgh, UK: Edinburg University Press, pp. 463-502

McClelland, J.L., Rumelhart, D.E. & the PDP Research Group (1986), *Parallel Distributed Processing. Vol. 2: Psychological and Biological Models*, Cambridge, MA: The MIT Press

Newell, A. & Simon, H.A. (1976), "Computer Science as Empirical Enquiry: Symbols and Search", *Communications of the Association for Computing Machinery*, **19**(3), pp. 113-126

Noë, A. (2004), *Action in Perception*, Cambridge, MA: The MIT Press

Nolfi, S. & Floreano, D. (2000), *Evolutionary Robotics*, Cambridge, MA: The MIT Press

Pfeifer, R. (1996), "Building 'Fungus Eaters': Design Principles of Autonomous Agents", in: P. Maes *et al.* (eds.), *Proc. of the 4th Int. Conf. on the Simulation of Adaptive Behavior*, Cambridge, MA: The MIT Press, p. 3-12

Pfeifer, R. & Scheier, C. (1999), *Understanding Intelligence*, Cambridge, MA: The MIT Press

Thompson, E. (2005), "Sensorimotor subjectivity and the enactive approach to experience", *Phenomenology and the Cognitive Sciences*, **4**(4), pp. 407-427

Thompson, E. (2007), *Mind in Life*, Cambridge, MA: The MIT Press

Thompson, E. & Varela, F.J. (2001), "Radical embodiment: neural dynamics and consciousness", *Trends in Cognitive Sciences*, **5**(10), pp. 418-425

Torrance, S. (2005), "In search of the enactive: Introduction to special issue on enactive experience", *Phenomenology and the Cognitive Sciences*, **4**(4), pp. 357-368

van Gelder, T. (1998), "The dynamical hypothesis in cognitive science", *Behavioral and Brain Sciences*, **21**(5), pp. 615-665

Varela, F.J. (1976), "Not One, Not Two", *The Co-Evolution Quarterly*, **12**, pp. 62-67

Varela, F.J. & Shear, J. (1999), 'First-person Methodologies: What, Why, How?', *Journal of Consciousness Studies*, **6**(2-3), pp. 1-14

Varela, F.J, Thompson, E. & Rosch, E. (1991), *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: The MIT Press

Weber, A. & Varela, F.J. (2002), "Life after Kant: Natural purposes and the autopoietic foundations of biological individuality", *Phenomenology and the Cognitive Sciences*, **1**, pp. 97-125

Wheeler, M. (2005), *Reconstructing the Cognitive World*, Cambridge, MA: The MIT Press

Ziemke, T. (2007), "What's life got to do with it?", in: A. Chella & R. Manzotti (eds.), *Artificial Consciousness*, Exeter, UK: Imprint Academic, pp. 48-66