Existentialism and Cognitive Science[*]

Michael Wheeler and Ezequiel Di Paolo

The Trailer

In the broadest possible terms, cognitive science is the multidisciplinary attempt to explain psychological phenomena in a wholly scientific manner. Exactly which disciplines count as members of the cognitive-scientific community remains, to some extent, an open question, partly because the mix of disciplines one thinks of as contributing to the overall project will ultimately reflect the specific theoretical outlook on mind, cognition and intelligence which one adopts. However, the interested bystander might typically glimpse some combination of artificial intelligence (including artificial life and certain areas of robotics), psychology (of various stripes), neuroscience, linguistics and philosophy. Traditionally, cognitive science has been dominated by the dual cognitivist principles of representationalism (intelligent systems work by building, storing and manipulating inner representations, where 'inner' standardly means 'realized in the brain') and computationalism (the processes by which those inner representations are built, stored and manipulated are computational in character). However, nonrepresentational and noncomputational approaches (e.g. some versions of the view that cognitive systems should be conceptualized as dynamical systems) are also part of the field's conceptual geography. (For a comprehensive philosophical introduction to cognitive science, see Clark 2001. For an unrivalled history, see Boden 2006.)

Now, according to one version of events, the story of the relationship between cognitive science and existentialism is rather like one of those Hollywood romances in which two people who start out hating each other, and who seem to be just about as ill-matched as anyone could possibly imagine, end up falling in love. In the present case, the happy couple still have some determined work to do before a discerning audience could be anywhere near confident that the match is one made to last, but, against the odds, there seems to be some genuine affection building, and who's to say how things might turn out. This is, we think, a story worth telling. In what follows, that's exactly what we shall endeavour to do.

Our story is best told by placing two historical plot lines alongside each other. The first begins with Hubert Dreyfus's influential critique of orthodox (i.e.,

[*] Wheeler, M. and Di Paolo, E. A. (2011). Existentialism and cognitive science, in Reynolds, J., Woodward, A., and Joseph, F. (eds) *The Continuum Companion to Existentialism*, Continuum, pp 241 – 259.

cognitivist) artificial intelligence, a critique driven predominantly by existentialist insights (see e.g Dreyfus 1990, 1992). If artificial intelligence (AI) is taken to be the intellectual core of cognitive science (as advocated by e.g. Boden 1990, p.1; 2006, chapter 4), then Dreyfus's critique generalizes straightforwardly from orthodox AI to orthodox cognitive science, and that's the way Dreyfus thinks of it. In a perhaps unexpected plot twist, however, Dreyfus's existentialism-driven onslaught is later transformed into a debate over how certain kinds of existentialist insight might be used productively to shape, mould and interpret research in (so-called) embodied cognitive science. (Cognitive science is embodied in form when it takes the details of the specific bodily structures and bodily manipulative capacities that a thinker enjoys, plus the ways in which those capacities interlock with particular external factors such as artefacts, to play an essential and transformative role in generating intelligent action and other psychological phenomena.)

As will become clear, the prequel to our first plot line would be the history of existentialist phenomenology, as represented by thinkers such as Heidegger (e.g. 1927) and Merleau-Ponty (e.g. 1962). What makes this envisaged prequel especially interesting for us is that, in an intriguing example of common intellectual descent, our second plot line starts precisely with one of Heidegger's students, Hans Jonas, and his existentialist phenomenology of life (Jonas 1966). Jonas's central insight, as expressed recently by Thompson, is that 'certain basic concepts needed to understand human experience turn out to be applicable to life itself', because 'certain existential structures of human life are an enriched version of those constitutive of all life' (Thompson 2007, 57). As our second plot line unfolds, we shall see that this existentially characterized deep continuity of life and mind, as revealed by Jonas, becomes one of the defining philosophical structures of one branch of (so-called) enactive cognitive science, an increasingly influential version of the embodied approach. So our two plot lines, with their closely related points of intellectual departure, will ultimately reconverge.

As indicated already, the channel through which existentialism and cognitive science began to take proper notice of each other was opened up by phenomenology. In its existentialist manifestation, phenomenology may paradigmatically be depicted as a theoretical (or perhaps meta-theoretical) philosophical enterprise that, through an attentive and sensitive examination of ordinary experience, aims to reveal the transcendental yet historical conditions which give that experience its form. Because these target structures are transcendentally presupposed by ordinary experience, they must in some sense be present with that experience, but they are not simply available to be read off

from its surface, hence the need for disciplined and careful phenomenological analysis to reveal them. And the historicality exhibited by the transcendental here is a consequence of (what, in this existentialist register, emerges as) the hermeneutic character of understanding in general, and thus of phenomenological understanding in particular. As an interpretative activity, phenomenological analysis is inevitably guided by certain historically embedded ways of thinking that the phenomenologist brings to the task, meaning that its results remain ceaselessly open to revision, enhancement and replacement. Beyond thinking of phenomenological analysis in this way, existentialist phenomenology is additionally conditioned by the characteristically existentialist conceptualization of human being as free and self-defining in (roughly) the following sense: as a human being, I am capable of transcending my own facticity. Here, 'my facticity' is understood as the physical, biological, psychological and historical features that might be established about me from the third-person perspective adopted by (among other explanatory practices) science. And transcendence is understood as the process of projection onto future possibilities in which I, in effect (although not necessarily reflectively), give value or meaning to those factical elements in terms of my projects and concerns, and thus bring forth a world of significance. Now, on the face of things, any research paradigm with this sort of profile is temperamentally bound to view cognitive science, which it is liable to interpret as being committed to an unobtainable-in-principle, objective scientific explanation of human being, with a good deal of intellectual suspicion. Intellectual suspicion is one thing, of course. It is altogether another to provide the kind of detailed critical indictment that the cognitive scientists themselves might actually take seriously. It is with this thought that our opening credits finally roll.

Dreyfus, Phenomenology and the Problem of Relevance

Psychologically and behaviourally, human beings are extraordinarily proficient at homing in on what is contextually relevant in a situation, while ignoring what is contextually irrelevant. This remains true, even in the sort of dynamically shifting and open-ended scenarios in which we often find ourselves. In short, human beings display a remarkable (although often unremarked upon) capacity to think and act in ways that are fluidly and flexibly sensitive to context-dependent relevance. Among many other things, a truly successful cognitive science would need to explain this capacity, and do so in a wholly scientific manner (i.e., without appeal to some magical, naturalistically undischarged relevance detector). In cognitive-scientific circles, this explanatory challenge is

sometimes known as the frame problem. We shall refer to it as the problem of relevance.

Viewed through the lens of an unreconstructed orthodox representational-computational cognitive science, the problem of relevance presents itself as the dual problem of (i) how to retrieve just those behaviour-guiding internal representations that are contextually appropriate and (ii) how to update those representations in contextually appropriate ways. The natural thought, given the lens through which we are currently looking, is that (i) and (ii) can be achieved if the intelligent agent specifies and tracks relevance, by systematically internally representing the key features of the contexts in which she finds herself. These context-specifying inner representations will in turn determine which first-order inner representations are relevant and so should be pressed into behaviour-guiding service. This might seem like an intuitively promising strategy. However, with the influence of existentialist phenomenologists such as Heidegger and Merleau-Ponty firmly in the foreground, Hubert Dreyfus has argued that, ultimately, it must fail. (For a full explication of the intertwined considerations that we are about to summarize, see e.g. Dreyfus 1990; for further analysis and discussion, see e.g. Dreyfus 2008, Wheeler 2005, 2008, 2010b, Cappuccio and Wheeler 2010, Rietveld forthcoming.)

Dreyfus's critique has three strands. First, Heideggerian phenomenological analysis reveals contexts to be complex, network-like semantic structures defined with reference to the concerns and projects (or projections – see above) of human agents (see e.g Heidegger 1927, p.116). For example, my laptop is currently involved in, or, as one might say, affords, an act of text-editing; that text-editing is involved in writing a document; that document-writing is involved in meeting a professional deadline; and that meeting of a professional deadline is involved in (it is done for the sake of) my project of being a good academic. But phenomenology discloses human activity as sensitive not only to what Rietveld calls the 'figure-affordance we are currently directed at and responding to' but also to what he calls 'a multiplicity of more marginally present ground-affordances that solicit us as well' (Rietveld forthcoming, 6). So the context-determining links to which my activity is currently sensitive, either actively or potentially (where 'potentially' signals the presence of a certain priming for attention rather than the mere possibility of relevance – see below), might be traced not only, in the active register, from laptops, to text editing, to document-writing, to professional deadlines, to the project of being a good academic, but also, in the register of potentiality, from the post-it note reminder stuck to the laptop screen, to the need to buy milk and bread on the way home, to the project of being a good partner and father. In this way, contexts spread

out, embed, overlap and combine to form the diffuse webs of relevance-determining relations that Heidegger (e.g. 1927, 118) once called totalities of involvements. According to Dreyfus, this has an important implication: because individual contexts inevitably leak into these massively holistic structures, they resist any determinate specification in the manner demanded by the orthodox representationalist strategy.

Secondly, Dreyfus interprets our fluid and flexible capacity for responding to relevance as at root a skill, understood as a form of knowing-how. In the background here is Heidegger's concept of circumspection. 'Circumspection' is Heidegger's term for (roughly) the adaptive sensitivity to context exhibited by our everyday skilled practical activity, a phenomenon which he identifies as the distinctive 'kind of sight' that action possesses (Heidegger 1927, 99). Building on this idea of a distinctive kind of sight or knowledge, Dreyfus claims that the sort of skilled know-how at work in human sensitivity to relevance cannot be reduced to, and thus cannot be exhaustively accounted for by, the kind of knowledge-that-something-is-the-case paradigmatically associated with representational content.

Finally, Dreyfus predicts that, and explains why, a vicious regress will accompany any attempt to specify relevance through the introduction of inner representations whose function is to bind context-dependent features to entities. According to Dreyfus, Any such second-order representational structures will need to have their own contextual relevance specified by third-order representations. But these new third-order structures will need to have their contextual relevance specified by fourth-order representations, and so on. One driver for this analysis is Heidegger's somewhat sketchy treatment of what he calls value-predicates (in effect Dreyfus's representations of context-dependent features; see Heidegger 1927, 132; for discussion, see Dreyfus 1990, Wheeler 2005). Heidegger claims that adding value-predicates to context-independent primitives (e.g. raw sense data or, to give the argument a more contemporary tone, light-intensity gradients at the retina) can never be the ultimate source of relevance, since each such value-predicate requires further structures of the same kind to determine its contextual relevance.

On one reading of the foregoing set of considerations, Dreyfus's existentialism-influenced message is that representations cannot solve the problem of relevance. However, Dreyfus goes further, by suggesting that, from the perspective of existentialist phenomenology, the problem of relevance is, at least partly, an artefact of representationalism. As he put it recently, 'for Heidegger, all representational accounts are part of the problem' (Dreyfus 2008, 358). Here

there is an important subtlety to bring out. As Dreyfus (2008) has made clear, representations may sometimes figure in the phenomenology and the neural underpinning of skilled know-how, since while some cases of skilled know-how are cases of (what he calls) absorbed coping, in which representations play no part (Heidegger's domain of readiness-to-hand), others are cases in which absorbed coping breaks down and the agent confronts a context-embedded problem to solve (e.g. the failure of the laptop's Internet connection that requires repair so that the activity of document preparation can continue; Heidegger's domain of unreadiness-to-hand). In cases like the latter, inner representations may form part of the cognitive response. For Dreyfus, however, the phenomenon that ultimately explains the relevance-sensitivity of human action, and thus neutralizes the problem of relevance, is ontologically more basic than nonrepresentational ready-to-hand coping or representational practical problem solving. He dubs that phenomenon background coping, understood as a nonrepresentational knowing how to get around one's world. As he puts it (Dreyfus 2008, 345-6), the 'all coping, including unready-to-hand coping, takes place on the background of [a] basic nonrepresentational, holistic, absorbed, kind of intentionality, which Heidegger calls being-in-the-world'.

Despite Dreyfus's talk of background coping being a species of intentionality that is more fundamental than skillful coping, it should not be thought of as a wholly separate phenomenon (Cappuccio and Wheeler 2010). Rather, as Merleau-Ponty (1962, 159) points out, 'movement and background are, in fact, only artificially separated stages of a unique totality'. Nevertheless, it is at the level of background coping that, to borrow an example from Gallagher (2008), the skilled climber's know-how first opens up the world as a familiar place of climbable mountains. When poised to engage in the action of climbing a mountain, the skilled climber does not build an inner representation of the mountain and infer from that plus additionally represented knowledge of her own abilities that it is climbable by her. Rather, from a certain distance, in particular visual conditions, the mountain 'simply' looks climbable to her. Her climbing know-how is 'sedimented' in how the mountain looks to her. Background coping may thus be illuminated by Merleau-Ponty's (1962) notion of the intentional arc, according to which skills are not internally represented, but are realized as contextually situated solicitations by one's environment that tend to become more fine grained with experience (cf. Dreyfus 2008, 340). Rietveld (forthcoming) provides a fuller phenomenological picture of background coping understood in terms of Merleau-Pontian solicitations, by drawing a distinction (referred to earlier) between different kinds of affordance (possibilities for action presented by the environment). Given a specific situation, some affordances are mere possibilities for action, where 'mere' signals

the fact that although the agent could respond to them in some way, such a response would be contextually inappropriate. In the same situation, however, some affordances, precisely because they are either directly contextually relevant to the present task at hand, or have proved to be relevant in similar situations in the past, prime us for action and thus, as Rietveld (forthcoming, 5) puts it, render us ready to act in appropriate ways by being bodily potentiating. Affordances of the latter kind are identified by Rietveld as a solicitations, divided into figure (relevant) affordances (those with which we are actively concerned) and ground (relevant) affordances (those with which we are not currently concerned but for which we are currently potentiated, and which are thus poised to summon us to act) (see Rietveld forthcoming, 5-9).

Although Rietveld doesn't put things in quite the way that we are about to, his analysis of background coping suggests that acts of transcendence – concrete instances of projection onto possibilities (see above) – need to be understood more specifically as acts of projection onto relevant possibilities, interpreted as embodied potentiations or solicitations. The existentialist challenge to cognitive science to explain transcendence is thus fully revealed as being to show how naturalistically unmysterious states and mechanisms may causally underpin background coping, understood in terms of solicitations. At this point it is worth stressing that although Dreyfus is sometimes attributed with the view that artificial intelligence, and by extension cognitive science, is impossible, this is to mis-state his position, which more accurately is that cognitive science as we (mostly) know it falls short of explaining human behaviour, with that shortfall explained in large part by the field's adherence to representationalism. So are there any cognitive-scientific models out there that might conceivably satisfy the Dreyfusian phenomenologist and the cognitive scientist? Perhaps there are. In recent work, Dreyfus (e.g. 2008) has cited with approval the neurodynamical framework developed by Freeman (2000), in which the brain is conceptualized as a nonrepresentational dynamical system primed by past experience to actively pick up and enrich significance, a system whose constantly shifting attractor landscape is identified as physically grounding Merleau-Ponty's intentional arc by causally explaining how newly encountered significances may interact with existing patterns of inner organization to create new global structures for interpreting and responding to stimuli.

We have just shifted philosophical key. The emerging idea is that existentialist phenomenology might have a positive role to play in revealing phenomena and processes that cognitive science might profitably explore – indeed, that existentialist phenomenology might even become a member of the cognitive-scientific community and benefit from a collaborative engagement with the

latter. This idea has also been explored by Wheeler (2005, 2008, 2010b) in his development and defence of (what he identifies explicitly as) a Heideggerian embodied cognitive science.

(Note: For further examples, not mentioned elsewhere in this piece, in which existentialist phenomenology has been used to make a positive contribution to embodied cognitive science, see e.g. Gallagher 2005, Kiverstein forthcoming, and Rowlands 2010, among others. For an innovative empirical study in embodied cognition that seeks to isolate the psychological signature of Heidegger's distinction between readiness-to-hand and unreadiness-to-hand, see Dotov et al. 2010. For recent arguments which conclude that the project of positive integration will be thwarted, unless cognitive science can divest itself of the kind of naturalism that Wheeler, for example, takes to be at the philosophical heart of the field, see Ratcliffe forthcoming, Rehberg forthcoming.)

Like Dreyfus, Wheeler takes the problem of relevance to be a central challenge for cognitive science. His analysis differs from Dreyfus's, however, in drawing a distinction between two different dimensions to the problem of relevance, the intra-context problem, which challenges us to say how a naturalistically discharged system is able to achieve appropriate flexible and fluid action within a context, and the inter-context problem, which challenges us to say how a naturalistically discharged system is able to flexibly and fluidly switch between contexts in a relevance-sensitive manner. According to Wheeler, the intra-context problem of relevance may be solved by what he calls special-purpose adaptive couplings. His favourite example is drawn from the domain of insect behaviour, which might set some alarm bells ringing in existentialist circles (and elsewhere), but the genuine differences between insect and human behaviour (not least in terms of the complexity of the contexts in which that behaviour is embedded) should not blind us to the fact that the context-sensitivity in question may be causally achieved by similar underlying mechanisms (for discussion, see Wheeler 2008; see also below on the nature of animality). The example, then, concerns the female cricket's capacity to track a species-specific auditory advertisement produced by the male. Robotic modelling by Webb (1994) suggests that this is achieved through a combination of the basic anatomical structure of the female cricket's peripheral auditory system (which ensures that the amplitude of her ear-drum vibration will be higher on the side closer to a sound-source) and the activation profiles of two interneurons that are tightly coupled with the specific temporal pattern of the male's song (such that only signals with the right temporal pattern will result in the female turning towards the sound source). Because this mechanism works correctly only in the

presence of the right, contextually relevant input, context is not something that must be reconstructed by the mechanism once it is activated. Rather, context is something that is always automatically present in that mechanism at the point of triggering. Wheeler (2008) interprets this as a kind of intrinsic context sensitivity that solves (or rather dissolves) the intra-context problem of relevance. So now what about the inter-context problem of relevance? Here is one possible model: fluid context-switching involves relevance-sensitive transitions between special-purpose adaptive couplings. It's here that Wheeler finds a place for the sort of shifting nonrepresentational dynamical system explored by Freeman and endorsed by Dreyfus. Such systems support a capacity for large-scale holistic reconfiguration that seems ripe to explain how a system could self-organise so as to realise different sets of special-purpose adaptive couplings.

Given Merleau-Ponty's point (see above) that movement and background are only artificially separated, Wheeler's position, as just sketched, results in a disagreement with Dreyfus over the cognitive science of background coping. For Wheeler, but not for Dreyfus, the mechanistic basis of background coping has the dual character of Freeman-style dynamics and situated special-purpose adaptive coupling – the different species of mechanism that, according to Wheeler, might explain inter-context sensitivity to relevance and intra-context sensitivity to relevance respectively. This dispute remains to be settled (although see Rietveld forthcoming for a discussion that finds in favour of Dreyfus). Whichever way it is resolved, however, the cognitive-scientific story here is incomplete, because the capacity for fluid systemic reorganization that arguably plays a key role in context-switching does not guarantee that the special-purpose adaptive couplings thereby brought on line (Wheeler) or the holistic reconfigurations of the system that transform its global dynamics (Dreyfus) will be the right (i.e. the newly contextually relevant) ones. All that is assured is that the system is a platform for the kind of flexibility that, when harnessed appropriately (i.e. in context-tracking ways), may help to generate fluid context-switching (for discussion, see Wheeler 2010b).

We have seen how the very same existentialist insights that shaped Dreyfus's critique of orthodox cognitive science are now helping in the development of a new kind of cognitive science. Interestingly, this sort of positive influence has been exerted via a different, although intimately related, channel.

Hans Jonas and the Enactive Approach to Life and Mind

In his introductory essay for a collection of his own papers translated from English into Spanish, Francisco Varela refers to his late discovery of Hans Jonas. He remarks on the surprising convergence of Jonas's philosophy with his own latest research directions (Varela 2000). It is no small homage to Jonas that the title Varela chose for this collection was "El fenómeno de la vida", a literal translation of the title of Jonas's own collection of edited essays (written in the 1950's and early 1960's) sketching his ambitious philosophy of life from an existentialist, yet scientifically informed, perspective (Jonas 1966). Varela's later work has been an attempt to engage with Jonas's bio-philosophy in an explicit dialogue with his own approach to life and cognition (Weber and Varela 2002), exploring what Thompson (2007) and others have described as the deep continuity between life and mind.

It is curious, though not unheard of, for converging ideas to make an independent appearance (simultaneously or separated in time) in rather remote regions of human learning, represented in this case by an existentialist philosopher with an interest in ethics and theology and an unconventional biologist and neuroscientist interested in non-Cartesian approaches to the mind. What might have been a minor scholarly curiosity turned out to be, in fact, a productive wellspring of novel thinking. Often radical and controversial, the ideas that originate from this encounter are at the core of an important line of theorizing within embodied approaches in cognitive science around which much discussion has been generated.

Let us first focus in some detail on Jonas's existential bio-philosophy. The lack of philosophical attention devoted to the roots of human existence and experience in organic life demands some examination. Even today, as discourses on bio-politics, bio-ethics and the precariousness of life take centre stage, current interest lies less in making explicit the connections between life, values and existence than in highlighting the technological subjection of biological substrates to new forms of control or in measuring human beings against other life forms. Presumably, the latter ends would benefit from the former. It is in this context that Hans Jonas's 'existential interpretation of biological fact' stands out. While affinities may be found with thinkers like Kurt Goldstein, Helmut Plessner, Georges Canguilhelm and others, Jonas arguably provides a unique handle on a thorny issue: the problem of why existence should be accompanied by any form of interiority and caring at all.

In effect, this is the same worry that drove Dreyfus to his Heideggerian critique of AI, or rather its obverse side. Accordingly, norms cannot be captured in an artificial system without a real embeddedness in a world of significance (cultural and socially mediated in the human case). Jonas's enquiries are attempts to dig further into the question of why this embeddedness would happen at all, what kind of conditions must be put in place for anything to count as a world in the first place. To answer this question unavoidably implies, at least in hindsight, examining the sort of entity that can qualify as having some form of self-concerned existence. Jonas constructs his first move through the unorthodox pairing of two contrasting forms of thought: Darwinism and phenomenology. The tension thus created is not resolved. It is instead used as a springboard for a bold proposal: all forms of life, even the simplest, have interiority and they all have a world.

In building a bridge that connects the human organism with the evolution of life on Earth, the supposed triumph of materialism (we are nothing but the result of the selective accumulation of random changes in chemical processes) presents us with 'the germ of its own overcoming' (Jonas 1966, 53). The reason for this is that our own experience as concernful embodied beings with an interior life is not denied by arguments of continuity with a world of efficient causes. On the contrary, both the experience and the arguments direct us to the phenomenon of life itself as a good place to seek the roots of what is often claimed to be a unique human privilege. The Darwinian bridge, under this view, turns out to be a two-way street. It is our own living experience that allows us to know life it is full reality, according to Jonas. 'Only life can know life' is the evocative slogan that sums up this view; full knowledge of life is not to be achieved unless we acknowledge our own insider's perspective on the topic.

If we accept as plausible that the experience of concern is not exclusively human (though it may have some specific characteristics in humans) and that all other physical living beings may also be, rather than appear, intrinsically teleological and in possession of an inner life, is this because they are living or simply because they are physical? For Jonas, it is a question of selecting the most informative option, the one that is more revealing. For Whitehead's philosophy of the organism, Jonas argues (1966, 95–96; 1968, 235, 241), there is no useful concept of challenge to organic identity since this kind of identity is in this view extended to cover all cases of physical identity, even that of particles that simply endure. Yet it seems a pragmatically vacuous extension of vocabulary to say that atoms die or molecules get sick. Most versions of pan-psychism are thus discarded since such precariousness is given to organisms by their singular mode

of identity: not the identity of inert permanence (A = A) but that of a dynamical form made of an ever-changing material substrate.

The break with the substantial mode of identity is achieved in *metabolism*, a self-affirming precarious process of constant regeneration of form within a flux of matter and energy. This is a feature of all life. For Jonas, this level of physical organization seems to have the necessary existential credentials:

> 1. the establishment of a distinct 'self' for which being is its own achievement and with organisational distinctions between inside and outside;
> 2. a precarious entity which is in constant environmental challenge, in need of material turnover and with the freedom to achieve it by regulating its exchanges with the environment; and
> 3. the establishment of norms following the logic of metabolism according to which otherwise neutral events, both internal and external, can be good or bad for the continuation of the organism.

Jonas's proposal is that metabolism is intrinsically teleological and all life possesses an inward dimension, a statement that cannot be arrived at by the unprepared, disembodied observer. Without our own inner experience as unquestionable datum, this proposal would be at most regulative, providing some help to the student of nature but in itself not derivable by reason, as indicated by the Kantian analysis of the intrinsic teleology and self-organisation of organisms in his Critique of Judgment (Kant 1790). The fact that metabolism sustains a dynamic form of identity (not coinciding with its material constitution at any given time except at the time of death) provides the possibility for the organism to become free. This freedom is expressed in the capability of the organism to engage with its medium in terms of the significance of a situation, thus contributing to its continuing dynamical autonomy and even opening up the possibility of novel value-making. However, this freedom is permitted by the meeting of very strict and specific material needs. It is a needful freedom. Rather than being paradoxical, this concept of freedom avoids the problems posed by determinism (and not solved by the inclusion of randomness) by operating on the relation of mediation between the self-sustaining, constantly becoming, identity and the 'target' of its worldly engagements. In this sense, the mode of realization of an autonomous process of identity-generation (like metabolism) establishes the sort of access this identity has to the norms that describe its different modes of viability. This access may be less or more mediated (the difference, say, between reacting with aversion to contact with a hot surface and planning our movements so as to avoid touching

it). Jonas's contention is that in the history of life and mind novel forms of increasingly mediated engagements have appeared allowing for more freedom at the cost of more precariousness.

Animals provide a clear example of such transitions. A new order of norms and values is founded in animality with the advent of self-generated motility and the co-emergence of perception, action and emotion. By putting a distance and a lapse between the tensions of need and the consummation of satisfaction, the temporality of the inner life is spatialised. Animals can appreciate right now the danger that is impinging on them from a distance. The future event becomes a distant but present possibility. This is the origin of a special relation with the world, that of perception and action, which is charged with internal significance, and hence with the development of an emotional dimension (what might have been an inner life of just need and satisfaction now becomes rich in possibilities such as fear, desire, apprehension, distension, tiredness, curiosity, etc.). But this comes at a cost of more severe energetic demands (allowing the necessary fast and continuing movements across varying environmental conditions without replenishment for long periods) and novel forms of risk.

As an example of how mediation enables new forms of freedom, consider the behaviour of several species of insects, like the water boatman, that are able to breathe underwater by trapping air bubbles (plastrons) using tiny hairs in the abdomen. The bubbles refill with oxygen due to the differences in partial pressure provoked by respiration and are prevented from collapsing by the hairs, thus potentially working indefinitely (see Turner 2000). These external lungs provide access to longer periods underwater thanks to a mediated regulation of environmental coupling (which is nevertheless riskier than normal breathing). The mediation in cases like this is so intimately connected with vital functions that the living system itself might be called extended. The issue at play in such reliable and conserved forms of mediation is, in each case, the question of the identity of such extended systems. New forms of life are built not so much 'on top' of existing ones but as possibilities for new forms of mediation and transformation of the relations between self-sustained identity and world.

Jonas recognizes other such transitions in modes of mediation in the history of life and mind, such as for instance those afforded by a complex visual system or the capacity to make images that leads to the birth of eidetic human projects and the distinction between truth and falsehood. It is doubtful whether any intrinsic gain is implied at the metabolic level by expanding the realm of freedom at the cost of increased precariousness. As Jonas points out, 'the

survival standard is inadequate for an evaluation of life' (Jonas 1966, p. 106). He goes on:

> It is one of the paradoxes of life that it employs means which modify the end and themselves become part of it. The feeling animal strives to preserve itself as a feeling, not just a metabolizing entity, i.e., it strives to continue the very activity of feeling: the perceiving animal strives to preserve itself as a perceiving entity—and so on. Without these faculties there would be much less to preserve, and this less of what is to be preserved is the same as the less wherewith it is preserved (ibid).

Effectively, such transitions in mediacy inaugurate a domain that feeds back on itself; they imply a new form of life. Not just in a metaphorical sense, but in the strict sense of a novel process of identity generation under-determined by metabolism.

The ideas in this landscape painted with broad strokes by Jonas are quite compelling and ripe for further exploration using the tools of systemic thinking and phenomenology. We can summarise the ideas that have the most direct and radical implications for cognitive science:

1. The use of a concept of identity whereby an individual is self-constructed by maintaining its own form in dynamic precarious conditions.
2. The implication from this form of identity that a living entity must thereby relate to the world with a specific interests and norms, i.e., the implication of an interior point of view.
3. The dialectics between living identity and the mediacy of its relation to the world leading to new forms of life of increased freedom and precariousness.

In cognitive science, the adoption of these ideas implies a radical break from traditional cognitivism (what we have earlier characterised as the dual principles of representationalism and computationalism) and possibly from other forms of functionalism as well. In contrast to 1., cognitivism does not have a theory of identity; the identity of a cognitive system is defined by convention or intuitive common-sense. In contrast to 2., cognitivism not only does not provide a good account of the origins of norms and values (as we have seen already), it also fails to see that such an account must inevitably involve the organisation and identity of the cognitive system; in traditional cognitive terms, how an agent is organised, what it is, how it should behave, what it does and what it cares about,

are all elements external to each other and brought together by a designer or an observer. And finally, in contrast to 3., cognitivism's way of understanding increasingly complex forms of mind is to measure their intuitive distance to the capabilities of an adult human being, as opposed to having a non-chauvinist method for understanding what is involved in the simultaneous transition to a new form of life and a new form of mind through the work of mediation.

Let us turn to how some of these ideas have influenced embodied approaches to cognition concerned with the deep continuity between life and mind.
Toward the last decade of his life, Francisco Varela explored a line of argument linking his early work on the autonomy of living systems with new research directions on embodied cognition (Varela et al 1991, Varela 1991, 1997). One important lesson from his early work with Humberto Maturana on the theory of autopoiesis (Maturana and Varela 1980) was the reclaiming of the living organism as a well-defined term for scientific discourse and as a proper level for the analysis of biological and cognitive phenomena. Science in general is comfortable at levels of explanation below the organism (genes, brain patterns, drives) or above it (environmental triggers, selection history, social structures), but rarely do slippery terms like 'individual', 'subject' or 'organism', let alone 'experience', play anything more than intuitive role in scientific discourse. The theory of autopoiesis is an attempt to propose a definition of a living system in such a way that the term would articulate a series of useful implications for the scientist and, therefore, would become a practical tool for scientific usage – an objective that has not quite been achieved, which is a topic for a different discussion. The declared goal of this theory is to examine the logical relations between two questions: what is the organisation of a living system and what are the possible ways in which a living system can relate to its world given this organisation.

Varela felt that the more pressing issues in this endeavour had not been fully examined in the original theory. These include issues such as the natural purposefulness of organisms, whether their teleology is real or merely an ascription by the observer, the organism's relation to the world in terms of significance, the origin of the norms that guide its behaviour, and so on. He addressed these issues following a systemic approach (Varela 1991, 1997): perhaps the purposefulness and sense-making of organisms are consequences of their organisation as self-producing autonomous systems. This has led to the proposal that it is indeed the living organisation that is responsible for the organism's capability to evaluate its encounters with the world. Sugar might be one of the many chemicals that we can observe surrounding a bacterium, but for the bacterium it is not a neutral presence. The value of sugar is manifested

behaviourally by a biased swimming up the sugar gradient with its consequences for the continued conservation of life. In Varela's words, encounters with the world are not neutral for an organism; they are invested with a 'surplus of signification' as a consequence of their self-producing nature.

A refinement of this argument followed Varela's encounter with Jonas's work (Weber and Varela 2002). For Varela, the element that Jonas's was lacking was a proper systemic approach to defining metabolism using systemic tools and the concept of self-organisation; a framework like the theory of autopoiesis, in short. For the enactive project, Jonas provides a rough map and some tools to navigate an immense landscape connecting various forms of life and mind, including those of human beings.

In casting Jonas's ideas in the language of systems science, Weber and Varela set on a road of continued conceptual refinement that is still transited today. As an example, the attempt to derive sense-making (the organism's capacity of relating to the world in terms of meaning, norms, and values) from simple autopoiesis (the ongoing self-construction of the organism) actually fails in its first instance. The reason for this is simple: if autopoiesis is all that is needed for a living system to be able to relate to its world in meaningful terms, then how are we to account for the graded nature of this relation, the fact that some things are appreciated by an organism as better than others, some risks are worth taking while others are not, some days as more full of struggle while others are more comfortable and relaxed? Across all of these graded differences the organism remains indistinctively alive; its autopoiesis does not change. Something else apart from an organisation that establishes an all-or-nothing distinction between life and death is needed for sense-making if this graded nature is to be explained. This extra characteristic is adaptivity (Di Paolo 2005), in short: a capacity that the organism has, in some cases, to revert the tendencies that, if allowed to continue, would result in its death. With this capacity (which comes in a large variety of forms and may be transformed during the organism's lifetime), it is possible to recover both the graded nature of our experience in making sense of the world as well as the spirit of Varela's starting intuitions. The refined argument now reads: sense-making implies both the presence of a self-sustained precarious organisation (like autopoiesis) and some form of adaptivity.

Jonas's key contributions are thus given a solid basis by the enactive approach (without implying that this endeavour is yet finished). This basis enables the conceptual articulation needed to examine several of the blind-spots of cognitivism, and this has led to a series of new proposals and critiques (see,

Stewart, Gapenne and Di Paolo 2010). For instance, enactive ideas have provided a new angle to debates on the extended mind hypothesis (Wheeler 2010a, Di Paolo 2009, Thompson and Stapleton 2009) where, enactivists argue, the concepts of autonomy, precariousness and sense-making elaborated above throw new light into how to determine what constitutes a cognitive system. Similar concerns have motivated more precise definitions of agency based on Jonasian arguments of continuity between life and mind (Barandiaran, Di Paolo and Rohde 2009). Computer models based on this approach to agency have provided insights on the relation between metabolism and behaviour in protocells and bacteria (Egbert and Di Paolo 2009, Egbert, Barandiaran and Di Paolo, 2010).

The concepts of autonomy and sense-making have been applied to a theory of social cognition less concerned with postulating mentalizing capabilities for understanding others' mental states and more focused on the processes of embodied interaction and participatory understanding (De Jaegher and Di Paolo 2007, De Jaegher 2009, De Jaegher, Di Paolo and Gallagher 2010, Di Paolo, Rohde and Iizuka 2008, Fuchs and De Jaegher 2009, Froese and Di Paolo 2009, McGann and De Jaegher 2009). The concern with experience and identity alerted researchers to problems with otherwise embodied proposals, like the sensorimotor approach to perception and consciousness (O'Regan and Noë 2001). Thompson (2007) has critiqued this approach for lacking a proper place for the autonomy of the cognitive system, which is phenomenologically translated as a lack of a good account of the subjectivity of personal experience. Other offshoots of enactive thought include a non-representational perspective on mental imagery (Thompson 2007), neuro-phenomenological accounts of the dynamics of first-person time-consciousness (Varela 1999), the fine time-structure of neural self-organisation in perception (Varela et al 2001), elucidations of the role of goal-directness in action (McGann 2007), refinements to notions of skills and perceptual modalities (McGann 2010), and work on developmental robotics (Vernon 2010) and evolutionary robotics (Di Paolo and Iizuka 2008, Rohde 2010).

It may be too early to fully evaluate these new developments – many of which are still making their way into more mainstream regions of cognitive science. It is, however, already remarkable that they all seem to derive from the encounter between Varela and Jonas's existential bio-philosophy. It is as if Jonas's insights, precisely because they originate in concerns that are far removed from mainstream cognitive science, may have served to unblock some of the most resilient impasses the field has had to deal with over the last 50 years. Remarkably, the very same thing could be said in relation to the insights of

Heidegger and Merleau-Ponty, insights that, as we saw earlier, have been shaping recent cognitive-scientific approaches to the problem of relevance. Our two plot lines have finally reconverged.

References

Boden, M. A. (1990), 'Introduction'. In M. A. Boden (ed.). The Philosophy of Artificial Intelligence. Oxford: Oxford University Press.

Boden, M. A. (2006), Mind As Machine: A History of Cognitive Science (two volumes). Oxford: Oxford University Press.

Barandiaran, X., Di Paolo, E., and Rohde, M. (2009), 'Defining agency individuality, normativity, asymmetry and spatio-temporality in action'. Adaptive Behavior 17 (5): 367–386.

Cappuccio, M. and Wheeler, M. (2010), 'When the Twain Meet: Could the Study of Mind be a Meeting of Minds?', in . J. Reynolds, E. Mares, J. Williams and J. Chase (eds.), Postanalytic and Metacontinental: Crossing Philosophical Divides. London: Continuum.

Clark, A. (2001), Mindware: an Introduction to the Philosophy of Cognitive Science. Oxford: Oxford University Press.

De Jaegher, H. (2009), 'Social understanding through direct perception? Yes, by interacting'. Consciousness and Cognition 18 (2): 535–542.

De Jaegher, H., and Di Paolo, E. (2007), 'Participatory sense-making: An enactive approach to social cognition'. Phenomenology and the Cognitive Sciences 6 (4): 485–507.

Di Paolo, E. A. (2005), 'Autopoiesis, adaptivity, teleology, agency'. Phenomenology and the Cognitive Sciences 4:429–452.

Di Paolo, E. A. (2009), 'Extended life'. Topoi 28:9–21.

Di Paolo, E. A., and Iizuka, H. (2008), 'How (not) to model autonomous behaviour'. BioSystems 91:409-423.

Di Paolo, E., Rohde, M., and Iizuka, H. (2008), 'Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing'. New Ideas in Psychology 26 (2): 278–294.

Dotov, D.G., Nie, L., Chemero, A. (2010), A demonstration of the transition from ready-to-hand to unready-to-hand'. PLoS ONE 5(3): e9433. doi:10.1371/journal.pone.0009433

Dreyfus, H. L. (1990), Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I. Cambridge, Mass.: MIT Press.

Dreyfus, H. L., (1992), What Computers Still Can't Do: A Critique of Artificial Reason. Cambridge, Mass.: MIT Press.

Dreyfus, H. L., (2008) , Why Heideggerian AI failed and how fixing it would require making it more Heideggerian', in P. Husbands, O. Holland and M. Wheeler (eds.) The Mechanical Mind in History, Cambridge, Mass.: MIT Press, pp. 331–71. (A shortened version of this paper appears under the same title in Philosophical Psychology 20/2 (2007): 247–68. Another version appears under the same title in Artificial Intelligence 171 (2007): 1137–60. Page numbers refer to the Husbands et al. (eds.) version.)

Egbert, M. and Di Paolo, E. A. (2009), 'Integrating behavior and autopoiesis: An exploration in computational chemo-ethology'. Adaptive Behavior, 17: 387-401

Egbert, M., Barandiaran, X. and Di Paolo, E. A. (2010), 'A minimal model of metabolism-based chemotaxis', PLoS Computational Biology, 6(12): e1001004. doi:10.1371/journal.pcbi.1001004.

Freeman, W. (2000), How Brains Make Up Their Minds. New York: Columbia University Press.

Froese, T. and Di Paolo, E. A. (2009), 'Sociality and the life-mind continuity thesis'. Phenomenology and the Cognitive Sciences, 8(4), 439-463.

Fuchs, T and H De Jaegher (2009), 'Enactive intersubjectivity: Participatory sense-making and mutual incorporation'. Phenomenology and the Cognitive Sciences, 8(4), 465-486.

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

Gallagher, S. (2008), 'Are minimal representations still representations?', International Journal of Philosophical Studies, 16 (3): 351-69, special issue on Situated Cognition: Perspectives from Phenomenology and Science, M. Ratcliffe and S. Gallagher (eds.).

Heidegger, M. (1927), Being and Time, trans. J. Macquarrie and E. Robinson. Oxford: Basil Blackwell.

Jonas, H. (1966), The Phenomenon of life: Towards a Philosophical Biology. Evanston, IL: Northwestern University Press.

Jonas, H. (1968), 'Biological foundations of individuality'. International Philosophical Quarterly, 8: 231–251.

Kant, I. (1790), The Critique of Judgement, Trans. Meredith, J.C., Oxford: Oxford University Press (this edition first published in 1952).

Kiverstein, J. 'Subjectivity without a subject-object distinction?', in J. Kiverstein and M. Wheeler (eds.), Heidegger and Cognitive Science. Basingstoke: Palgrave Macmillan.

Maturana, H. and Varela, F. J. (1980), Autopoiesis and Cognition: the Realization of the Living. Dordrecht, Holland: D. Reidel Publishing.

McGann, M. (2007), 'Enactive theorists do it on purpose'. Phenomenology and the Cognitive Sciences, 6(4), 463-483.

McGann, M. (2010), 'Perceptual modalities: Modes of presentation or modes of action?', Journal of Consciousness Studies. 17: 72-94.

McGann, M., and De Jaegher, H. (2009). 'Self-other contingencies: Enacting social perception'. Phenomenology and the Cognitive Sciences 8 (4): 417–437.

Merleau-Ponty, M. (1962), Phenomenology of Perception, trans. C. Smith, London: Routledge.

O'Regan, J. K., and Noë, A. (2001), 'A sensorimotor account of vision and visual consciousness'. Behavioral and Brain Sciences 24 (5): 883–917.

Ratcliffe, M. (forthcoming), 'There can be no cognitive science of Dasein', in J. Kiverstein and M. Wheeler (eds.), Heidegger and Cognitive Science. Basingstoke: Palgrave Macmillan.

Rehberg, A. (forthcoming), 'Heidegger and cognitive science - aporetic reflections', in J. Kiverstein and M. Wheeler (eds.), Heidegger and Cognitive Science. Basingstoke: Palgrave Macmillan.

Rietveld, E. (forthcoming), 'Context-switching and responsiveness to real relevance', in J. Kiverstein and M. Wheeler (eds.), Heidegger and Cognitive Science. Basingstoke: Palgrave Macmillan.

Rohde, M. (2009), Enaction, Embodiment, Evolutionary Robotics. Simulation Models for a Post-Cognitivist Science of Mind. Amsterdam and Paris: Atlantis Press.

Rowlands, M. (2010), The New Science of the Mind: from Extended Mind to Embodied Phenomenology. Cambridge Mass.: MIT Press.

Stewart, J., Gapenne, O. and Di Paolo, E. A. (eds) (2010), Enaction: Towards a New Paradigm for Cognitive Science. Cambridge, Mass.: MIT Press.

Thompson, E. (2007), Mind in life: Biology, Phenomenology, and the Sciences of Mind. Cambridge, Mass.: Harvard University Press.

Thompson, E. and Stapleton, M. (2009). 'Making sense of sense-making: Reflections on enactive and extended mind theories', Topoi, 28: 23-30.

Turner, J. S (2000), The Extended Organism: The Physiology of Animal-Built Structures, Cambridge, Mass.: Harvard University Press.

Varela, F. J. (1991), 'Organism: A meshwork of selfless selves', in A. I. Tauber (ed.) Organism and the Origin of the Self. Netherlands: Kluwer Academic, pp. 79–107.

Varela, F. J. (1997), 'Patterns of life: Intertwining identity and cognition'. Brain and Cognition 34: 72–87.

Varela, F. J. (1999), 'The specious present: A neurophenomenology of time consciousness', in J. Petitot, F. J. Varela, B. Pachoud and J.-M. Roy (eds.), Naturalizing Phenomenology. Stanford, CA: Stanford University Press, pp. 266–314.

Varela, F. J. (2000), El Fenómeno de la Vida. Editorial Dolmen, Santiago de Chile.

Varela, F. J., Lachaux, J.-P., Rodriguez, E., and Matinerie, J. (2001). 'The brainweb: phase synchronization and large-scale integration'. Nature Reviews. Neuroscience 2: 229–230.

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

Vernon, D. (2010), 'Enaction as a conceptual framework for developmental cognitive robotics'. Paladyn Journal of Behavioral Robotics, 1(2): 89-98.

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

Webb, B. (1994), 'Robotic Experiments in Cricket Phonotaxis', in Cliff, D., Husbands, P., Meyer, J.-A., and Wilson, S.W. (eds), From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior, Cambridge, Mass.: MIT Press, 45-54.

Wheeler, M. (2005), Reconstructing The Cognitive World: The Next Step. Cambridge, Mass.: MIT Press.

Wheeler, M. (2008), 'Cognition in context: Phenomenology, situated robotics and the frame problem'. International Journal of Philosophical Studies, 16 (3): 323-49, special issue on Situated Cognition: Perspectives from Phenomenology and Science, eds. M. Ratcliffe and S. Gallagher.

Wheeler, M. (2010a), 'Minds, things, and materiality', in Renfrew C. and Malafouris L. (eds.), The Cognitive Life of Things: Recasting the Boundaries of the Mind, Cambridge: McDonald Institute for Archaeological Research Publications.

Wheeler, M. (2010b), 'The problem of representation', in Gallagher, S. and Schmicking D. (eds.), Handbook of Phenomenology and Cognitive Science. Dordrecht: Springer, pp.319-336.