

Mind in Life or Life in Mind? Making Sense of Deep Continuity

Michael Wheeler

1. The Deep Continuity Thesis

One of the many ground-breaking themes in Evan Thompson's rich and thought-provoking book *Mind in Life: Biology, Phenomenology, and the Sciences of Mind* is his distinctive development and defence of an idea that he calls the *deep continuity of life and mind* – henceforth just *deep continuity*. Thompson introduces this idea as follows: "life and mind share a set of basic organizational principles, and the organizational properties distinctive of mind are an enriched version of those fundamental to life. Mind is life-like and life is mind-like" (p.128).¹ In this initial characterization, deep continuity is (as Thompson notes) tantamount to what others (e.g. Godfrey-Smith 1994, Wheeler 1997) have called the *strong continuity thesis of life and mind*.² Thompson claims, however, that these other theorists, in concentrating on organizational, functional or behavioural properties, have ignored a crucial aspect of life-mind continuity, namely its *phenomenological* dimension. The corrective, then, which recruits an insight that Thompson traces back to the work of Hans Jonas (1966), is to recognize that "certain basic concepts needed to understand human experience turn out to be applicable to life itself" (p.129). Such concepts (more on which below) include needful freedom, self-transcendence, and immanent purposiveness. In other words, "certain existential structures of human life are an enriched version of those constitutive of all life" (p.157).

In what follows I shall offer an all-to-brief response to Thompson's account of deep continuity. The bulk of the paper will be devoted to a critical analysis of the key elements out of which Thompson constructs his account, an analysis during which I highlight and discuss a number of internal problems and unresolved issues for the view. That done, I shall argue that Thompson's specification of the relationship between autopoiesis, cognition and life has the unpalatable effect of

¹ Unless otherwise indicated, all page numbers refer to Thompson 2007.

² For example, Godfrey-Smith (1994, p.83) characterizes strong continuity as the view that "[life] and mind have a common abstract pattern or set of basic organizational principles... Mind is literally life-like".

closing off what is arguably the most plausible form of deep continuity. I shall end by sketching (but no more than sketching) the thought that the most plausible form of deep continuity may yet be established within the kind of generic conceptual framework to which Thompson doesn't exactly warm, a framework, that is, which holds that the fundamental structures of life and mind are to be understood in terms of notions such as information and representation.

2. Autonomy and Cognition

What is the precise content of Thompson's deep continuity thesis? To make progress on this question, we need to reflect on various theoretical concepts that Thompson deploys as he explicates his version of the so-called *enactive* approach to life and mind. For the most part, these concepts are drawn (sometimes in a revised form) from the theory of autopoiesis – that powerful, but difficult to unravel, web of ideas concerning life and mind on which Thompson's enactivism is built.³

Perhaps the place to begin is with the least specialized of the concepts at work, namely that of *self-organization*. A self-organizing system is one in which certain intrasystemic components, on the basis of purely local rules (i.e. without the direction of some global executive control process), interact with each other in nonlinear ways so as to produce the emergence and maintenance of structured global order. Self-organization is now recognized as being a widespread phenomenon. Regularly cited examples in the literature include the Belousov-Zhabotinsky chemical reaction, lasers, slime moulds, foraging by ants, and flocking behaviour in creatures such as birds. Now, although self-organization is commonplace, only some self-organizing systems are what Thompson (following Varela) calls *autonomous*. For a self-organizing system also to be an autonomous system, the constituent processes of that system must "(i) recursively depend on each other for their generation and their realization as a network, (ii) constitute the system as a unity in whatever domain they exist, and (iii) determine a domain of possible interactions with the environment" (p.44; citing Varela 1979, p.55). Put slightly differently, an autonomous system is a network of interdependent processes whose recurrent activity (a) produces and maintains the very boundary that determines the identity of that network as a unitary

³ See, especially, chapter 5 of *Mind in Life*. For the canonical presentation of autopoietic theory, see Maturana and Varela 1980. For other autopoiesis-based statements of the enactive approach, see e.g. Varela et al. 1991, di Paolo 2005.

system (such that what counts as inside the system and what counts as outside the system co-emerge as functions of that endogenous activity), and concurrently (b) defines the ways in which that system may encounter perturbations from what is outside it while maintaining its organization and thus its viability, that is, while not disintegrating. Where successive environmentally induced perturbations to the autonomous system trigger only state changes that remain within that system's bounds of viability (such that organization is not lost), phenomena (a) and (b) together produce a history of *structural coupling* between that autonomous system and its environment.

Enaction is the process by which significance or relevance is brought forth through structural coupling. The core idea here is that the process of autonomous organization, in establishing the distinction between the self-maintenance and the collapse of system as a unity, institutes a norm of survival, and thereby the significance or relevance of certain environmental perturbations as either leaving the system organizationally intact or resulting in its disintegration. However, the bare norm of survival to which I have just given expression generates only an impoverished kind of significance, because it is based on no more than the robustness or conservation of systemic organization in the face of environmental perturbation. The system either survives any perturbations it experiences, or it doesn't. As Di Paolo (2005) usefully puts it, the norm of survival in play so far is an all-or-nothing affair. What seems clear, however, is that domains of significance enacted on the basis of structural coupling are regularly more complex in nature, in that they are places "of valence, of attraction and repulsion, approach or escape" (p.158). To illustrate this with the kind of example that Thompson himself often uses (e.g. p.158), bacteria swim towards the area containing the greatest concentration of glucose molecules. Thus, as a consequence of the specific metabolically realized autonomy of the bacteria, glucose emerges as – is brought forth as – significant for those organisms *as food*. The normative structure enacted here is not an all-or-nothing affair, but rather a gradient of better or worse outcomes for the autonomous system. Navigating this gradient requires the system to be able to alter its behaviour in response to changes in its environment, so as to exhibit a sensitivity to differences between states. In some way (perhaps implicitly), the system monitors how it is doing with respect to maintaining its viability, and thus regulates its behaviour accordingly, in order to improve its situation. In other words, the autonomous system is now also an *adaptive* system. The enactive process of generating graded significance on the basis of adaptive autonomy is what Thompson, following others (e.g. Varela 1991, Di Paolo 2005), calls *sense-making*.

The next idea from Thompson's conceptual toolkit that we need in order to carry out our analysis of deep continuity is the aforementioned notion of *autopoiesis* (Maturana and Varela 1980). Put crudely, autopoiesis is autonomy plus materiality. More specifically, to be autopoietic, an autonomous system must, through its own endogenous self-organizing dynamics, produce and maintain a *material* (or physical) boundary which distinguishes that system as a *material* (or physical) unity in the space in which it exists. The paradigm example of an autopoietic system is the single living cell, understood as a network of chemical reactions that produces its own membrane. The autopoietic organization of the living cell is, as Thompson describes it, "autonomy in the biochemical domain" (p.44). The precise relationship between autopoiesis and autonomy will concern us later, but one thing is worth noting immediately: autopoiesis is sufficient for autonomy.⁴

So now, what does all this tell us about life, mind, and the relationship between them? Let's begin with the notion of cognition. Thompson explicitly defines cognition as sense-making. Thus he says "cognition, in the present context, means the activity of sense-making" (p.159). Given that autonomy and adaptivity are individually necessary and jointly sufficient for sense-making (see above), the identification of cognition with sense-making makes autonomy and adaptivity individually necessary and jointly sufficient for cognition. However, we need to know how to interpret the phrase "in the present context". Thompson's text strongly suggests that this qualification is there to highlight the fact that we are concerned, in the first instance, with cognition "in its minimal biological form" (p.159), where the minimal biological form of cognition is the first (simplest, most basic) instance of an enacted domain of graded significance in nature. Thus Thompson directs our attention towards what he calls "the natural roots of intentionality", where what is at stake is an "analogue of the phenomenological notion of the disclosure of the world", understood as activity that brings forth a domain of graded significance on the model of our glucose-sensitive adaptive bacteria (p.159).

⁴ Although this accurately reflects the way Thompson introduces the term 'autopoiesis', he later gravitates toward using it in a 'wide' sense, so as to include adaptivity (e.g. p.158). It will suit my purposes to retain the narrower usage, and that's what should be assumed in the discussion that follows.

In spite of this drive towards the minimal and the basic, one shouldn't be tempted to think that anything less than sense-making through adaptive autonomy will be adequate for genuine cognition. As Thompson puts it:

Cognition is behavior or conduct in relation to meaning and norms that the system itself enacts or brings forth on the basis of its autonomy. We have seen that sense-making requires more than minimal autopoiesis: it requires autopoiesis enhanced with a capacity for adaptivity... [A]utopoiesis plus adaptivity entails sense-making, which is cognition in its minimal biological form. (p.159)

The attentive reader will no doubt have noticed that Thompson here talks of enhancing minimal *autopoiesis* rather than, as one might have expected from the momentum of my exposition, enhancing minimal *autonomy*. If we are concerned with the *sufficient* conditions for cognition, this is not an overly significant variation. As we have seen already, autopoiesis is sufficient for autonomy, so autopoiesis plus adaptivity will be sufficient for sense-making and thus for cognition in its minimal biological form. And it is "on the basis of its autonomy" that an autopoietic system of the right kind (i.e. one whose autonomy is adaptive in form) is able to generate cognitive activity. However, if we are concerned with the *necessary* conditions for cognition, Thompson's use of the term 'autopoietic' at a point where we might have expected him to use 'autonomous' immediately raises the question of whether there could be a nonautopoietic autonomous system. In other words, it raises the question of whether, for Thompson, autopoiesis is necessary for autonomy.

It might seem that Thompson's answer to this question is blindingly obvious. Indeed, he states that to "qualify as autonomous... a system does not have to be autopoietic in the strict sense (a self-producing bounded molecular system)" (p.44) and that "there can be autonomous systems that are not autopoietic if their constituent processes exhibit organizational closure in their domain of operation" (pp.106-7). We can understand these statements better if we consider the examples Thompson gives of autonomous systems, which, alongside the living cell, include the immune system (p.49, p.65), the nervous system (pp.46-7, p.65), insect colonies (p.44, p.65) and animal societies (p.65). As we have seen, the living cell is an autopoietic system, but, as Thompson points out, these other autonomous systems are not. In each case the categorization of the system as autopoietic or not turns on the nature of (a) the relevant system-nonsystem boundary and (b) the domain in which the system in question exists, as a network of constituent processes that generates a unity. For the living cell, the

system-nonsystem boundary (the cell membrane) and the domain of existence (the biochemical) are material. This specific kind of dual materiality is definitional of autopoiesis. We can contrast this with, for example, the nervous system and insect colonies. For the nervous system, the fundamental logic of which is to couple sensing and movement, the relevant system-nonsystem boundary and the domain of existence are fixed at the level of behaviour and intentional action (p.49); and for insect colonies the system-nonsystem boundary and the domain of existence are fixed at the level of sociality and territoriality (p.44). The issue, then, seems to be cut and dried: Thompson's view is that autopoiesis is not necessary for autonomy.

There is, however, an exegetical snag. There are passages in *Mind in Life* which strongly suggest that the interpretation just given must be wrong, and that, for Thompson, autopoiesis *is* necessary for autonomy. Consider, for example, the following sequence of claims:

Agency and meaning require autonomy; minimal agency and meaning require minimal autonomy. Minimal autonomy *depends* on macromolecules but requires that those macromolecules be *organized* in a particular way, namely, in the autopoietic way. (p.160)

It is hard to read this passage without taking away the message that autopoiesis is necessary for autonomy. If autonomy, agency and meaning *in their minimal forms* require autopoiesis, then the most natural thought is that *all* forms of agency, meaning and autonomy do. After all, 'necessary for the minimal form of X' sounds achingly close to 'minimally necessary for X'. What breeds confidence that we must put any such thought aside *as an interpretation of Thompson* is that it clashes unhelpfully with his claim that the immune system, the nervous system, insect colonies and animal societies are autonomous systems. Recall the dual materiality of self-generated systemic boundary and domain of existence that, as we have seen, is definitional of autopoiesis. If autopoiesis is indeed necessary for autonomy, then the systems currently under consideration, by virtue of the natures of their self-generated systemic boundaries and domains of existence (see above), would not count as autonomous. The most prominent difficulty would then be that, if autonomy is necessary for sense-making and thus for cognition, the nervous system (no longer an autonomous system) could not be a basis for cognition. And that really would upset the applecart.

In the preceding paragraphs, I have suggested that there is an inconsistency in Thompson's treatment of the relationship between autopoiesis and autonomy, but that we can be confident that he would resolve that inconsistency in a particular way. Nevertheless, we should pause for a moment to wonder whether there is an alternative reading of the anomalous passage that would avoid plunging him into the difficulty first place. So let's try something out. If we unpack 'minimal' as 'minimal biological form of', in line with Thompson's analysis of cognition (see above), we might interpret him as holding that although the minimal biological form of autonomy, and thus the minimal biological forms of agency and meaning, require autopoiesis, more advanced forms of autonomy, agency and meaning are able to divest themselves of any such dependence.

Attributing this view to Thompson would avoid the troublesome inconsistency with which we are presently concerned (and, incidentally, explain why he is comfortable substituting 'autopoiesis' for 'autonomy' when characterizing cognition in its minimal biological form – see earlier discussion). That said, it would saddle him with a new burden, by focussing critical attention on the contribution made by the materiality requirements that distinguish autopoiesis as a phenomenon over and above autonomy. In identifying autopoiesis as necessary for minimal autonomy, Thompson's goal is to reveal "the ground from which the seeds of intentional action grow" (pp.160-1). But the only thing that autopoiesis adds to the concept of autonomy is the dual materiality of the systemic boundary and the systemic domain of existence. And it is genuinely hard to see what special substantive contribution is made to our explanation of the genesis of intentional action by recognizing that the minimal biological form of autonomy exhibits such dual materiality, given that such materiality is apparently expendable by the time that biology gets as far as a nervous system. After all, although the distinctive kind of materiality that autopoiesis signals is an interesting and important feature of the cell, the idea of minimal autonomy already provides for a self-producing unity capable of instantiating a history of structural coupling, and it is an enrichment of *that very capacity for structural coupling*, through the addition of adaptivity, that accounts finally for sense-making and thus for cognition. So the dual materiality of the systemic boundary and the domain of existence is not required for the realization of these phenomena.

In response to the preceding analysis it might be suggested that although the boundary and domain of existence of the autonomous system in question need not be material in the relevant sense, nevertheless there must be autopoietic

materiality somehow at the centre of any system that genuinely has its own perspective on the world.⁵ Of course, some account would need to be given of what 'at the centre' means here, but in any case I think the considerations that figured in the original analysis are very likely to continue to apply. Consider: As it happens, biological cells are not merely autonomous, they are autopoietic. But let's imagine that the physical building blocks of organisms were somehow autonomous but not autopoietic. Given that autonomy is sufficient to establish a self-producing unity, and that adaptiveness is a modulation of that organization, it seems that despite this assumed lack of any autopoietic dimension at the centre of things, cognition and intentionality would still enjoy an impeccable naturalistic grounding.

3. Autopoiesis and Life

After some clarificatory discussion, we now have a grip on Thompson's view of cognition: cognition is sense-making and so requires adaptive autonomy but not autopoiesis. Now, what about life? Thompson's striking claim is that life is autopoiesis plus cognition (p.158). Since cognition is sense-making as established by adaptive autonomy, this is formally equivalent to claiming that life is adaptive autopoiesis (adaptive autonomy in the biochemical domain). Thompson's development of this view of life is complicated, but in the present context we can concentrate on two things – the neo-Kantian role played by autopoiesis and the neo-Jonsonian role played by the combination of needful freedom and self-transcendence.

Thompson uses the concepts of self-organization and autopoiesis to give a contemporary reinterpretation of Kant's (1790) claim that organisms are *natural purposes* (pp.129-49). In brief, the picture is that organisms, like artefacts, are systems in which each component exists for the sake of the others in the context of the organized whole, in such a way that the cause-effect relations that characterize the workings of the system are also means-end relations. That means that teleological (purposive) language is needed to characterize the system. However, whereas in the case of artefacts the purposiveness in question is extrinsic (e.g., established by the pre-existing plan of a human designer), in the case of organisms the purposiveness is intrinsic or, as Thompson prefers to say, *immanent*. A property is immanent if it is neither a "nonrelational [and so unanalysable, see p.146] property of something internal to the system... nor a

⁵ Something like this thought seems to be suggested by Di Paolo's (2005) claim that autopoiesis "provides a self-distinct physical system that can be at the centre of a perspective on the world".

property determined by something outside the system... [but rather] a constitutive property the whole system possesses because of the way the system is organized" (p.146). The purposiveness of an organism is necessarily immanent because the means-ends relations that characterize its parts are established by the distinctive *self*-organizing and *self*-producing activity of the system according to which "each of its parts is both a product and a producer of the other parts, so that the system is a self-organizing whole" (pp.145-6). And Thompson takes it that if organisms were recognized to be autopoietic in nature, that would be sufficient to explain their distinctive self-organizing and self-producing character and thus the immanence of their purposiveness (p.146).

What are we to make of Thompson's account of immanent purposiveness? It is worth just noting that there is a dispute to be settled here over whether or not any multicellular organism is itself strictly autopoietic, as opposed to being built out of autopoietic elements (see pp.105-7 for discussion). Fortunately, however, having noted this issue, we can simply ignore it, since it seems open to Thompson to hold that even if multicellular organisms are not themselves strictly autopoietic, the immanent purposiveness of any multicellular organism may be inherited from the cells that together make it up, given that each of those cells will be an autopoietic system and thus will individually realize immanent purposiveness. So let's agree that Thompson can make good on the claim that autopoiesis is sufficient for immanent purposiveness. A further question immediately suggests itself: for Thompson, is autopoiesis necessary for immanent purposiveness? To answer this question we need to dig just below the surface of the text. During his summary of di Paolo's (2005) argument that sense-making requires adaptivity (see above), Thompson notes that, for di Paolo, "[a]daptivity needs to be established on the basis of autopoiesis; otherwise sense-making is not original to the system but merely attributed from the outside" (p.148). What drives this claim is the thought that, without the connection to the specific self-distinguishing process of autopoiesis, the *teleological structures* of sense-making would not be original to the activity of the system (that is, immanent), but would merely be attributable to the system by some external observer. The same dependency would presumably hold for the teleology appropriate to any bare (all-or-nothing) norm of survival based on conservation rather than adaptivity, although of course a minimally autopoietic, and thus nonadaptive, system would not be cognitive and so, for Thompson, could not count as being alive. If all this is right, then, on di Paolo's view, autopoiesis is not only sufficient for immanent purposiveness, it is necessary too. And given that Thompson greets di Paolo's argument with approval, I take that to be a view which Thompson shares.

At this point it is tempting to resurrect the kind of worry that I raised earlier in connection with the claim that autopoiesis is necessary for the genesis of intentionality, and ask exactly why, in addition to the self-maintaining profile of autonomy, the dual materiality of autopoiesis is necessary to establish the immanence of the relevant teleology. One response that Thompson might conceivably pursue here would be that it is the specific kind of materiality enshrined in autopoiesis (in which a system-defining material boundary is endogenously produced and maintained) that is needed to give content to the idea (introduced earlier) that each part of a system displaying immanent purposiveness must be both a product and a producer of the other parts. That would explain why autopoiesis, and not merely autonomy, is necessary for immanent purposiveness and thus for life. So how compelling is the idea that the kind of productive interdependence at issue requires autopoietic materiality? That depends largely on how the notion of a 'systemic part' is to be understood. As we have seen, within Thompson's framework, the domain of existence of an autonomous system – the ontological level at which its identity is fixed – need not be one of first-order materiality; it might, for example, be a domain of social relations rather than biochemistry. Presumably, then, there is a relevant notion of a 'systemic part' that is fixed at an ontological level other than first-order materiality. For example, a part of an animal social system may be fixed with respect to a level in a dominance hierarchy. But if this is right, then it is significant that Thompson's approach already contains the idea that the constituent processes of an autonomous system "recursively depend on each other for their generation and their realization as a network" (see above). For if, as seems likely, the constituent processes of a system count as systemic parts, or if, as also seems likely, the idea of recursive dependence stretches to systemic parts that ontologically speaking aren't processes (if any such parts there be), then it seems the notion of autonomy already gives us a picture in which each systemic part is both a product and a producer of the other systemic parts. And that is a picture of immanent purposiveness. But if autonomy is sufficient for immanent purposiveness, then even though autopoietic systems will exhibit immanent purposiveness in virtue of being autonomous systems, autopoiesis is not needed for immanent purposiveness.

Whatever problems Thompson faces in carving out a distinctive role for autopoiesis in explaining the phenomenon of life, the fundamental structure of his own final position is, I think, clear enough: cognition is sense-making, while life is autopoiesis plus cognition, which is tantamount to saying that cognition is adaptive autonomy, while life is adaptive autopoiesis. However, this structural

description of things, although accurate, does not do justice to the full richness of the cognitive dimension of life, as Thompson sees it. According to Thompson, living systems display the intertwined properties of *needful freedom* and *self-transcendence*. Both spring from the character of metabolism. Although, at any particular moment, an organism is realized by a finite collection of material elements, the distinctive form of the organism cannot be identified with that matter, precisely because metabolism involves a constant turnover of material elements. Life as a phenomenon thus exhibits a kind of *freedom from* material constitution, simultaneous with a dependence on, a *need for*, the material exchanges with the environment (roughly, eating and excretion) that metabolism encompasses. Self-transcendence emphasizes the fact that such metabolically grounded needful freedom may be interpreted in terms of a kind of projective organization in which organismic being involves a relentless going beyond of current state, a transcending (surpassing) of present self-identity. These properties of needful freedom and self-transcendence are explicitly introduced by Thompson as analogues of existential structures that may be revealed by a phenomenological analysis of conscious human experience. For example, and very roughly, according to Heidegger (1927), I confront every concrete situation in which I find myself as a range of possibilities for the future realization of some pattern of human being, meaning that human being essentially involves a projection beyond, a freedom from, and a surpassing of, my current concrete state. In this way, then, concepts that, from a phenomenological point of view, describe universal structural features of human conscious experience also serve to locate certain fundamental properties of any living system.

4. Two Kinds of Deep Continuity

Up to now we have been engaged in a critical exposition of the key conceptual elements out of which Thompson constructs his account of deep continuity. Our next task is to build on that exercise to ask precisely what deep continuity might amount to, for Thompson. There are no doubt many ways in which one might pursue these goals. I shall content myself with the following twofold strategy. I shall present two different interpretations of deep continuity that (the preceding analysis suggests) may be extracted from *Mind in Life*. Of each of these interpretations I shall ask (i) How does it conceive the enrichment process that deep continuity involves?, and (ii) Is it consistent with another view that is often mentioned alongside enactivism as part of a generic embodied cognition perspective, namely the *extended cognition hypothesis*? Let's pause momentarily to comment on each of these questions.

Regarding (i): According to Thompson's notion of deep continuity, the organizational and existential properties distinctive of mind are an enriched version of those fundamental to life (see above). One might be forgiven for taking this to mean that there are certain *noncognitive* properties of all living entities that, when enriched in specific ways, generate the phenomena of mind and cognition, phenomena that are exhibited only by a subset of living things. I think that this must be how Godfrey-Smith (1994) understands his own, related thesis of strong continuity (again, see above). It is certainly how I have always understood the idea (see Wheeler 1997). But, as we shall see, this seemingly natural interpretation of the enrichment process at the heart of deep continuity cannot be endorsed by Thompson. At root, this parting of the ways is driven by the fact that the alternative visions of life-mind continuity that will be placed on the table embrace more than one conception of the *kind* of property that gets enriched in the evolutionary and developmental paths that lead from single cells to conscious human life. Put crudely, the question is whether or not what gets enriched is already mental in character *pre-enrichment*, so a blunt way to raise the right issues here would be to wonder whether it is possible for life and cognition to exist separately from each other. In other words, by the lights of some particular interpretation of deep continuity, could there be nonliving cognitive entities or, conversely, noncognitive living entities? That is how I shall approach question (i).

Regarding (ii): According to the extended cognition hypothesis (henceforth ExC), there are actual (in this world) cases of intelligent action in which thinking and thoughts – more precisely, the material vehicles that realize thinking and thoughts – are spatially distributed over brain, body and world, in such a way that the external (beyond-the-skin) factors concerned are rightly accorded cognitive status. In other words, the physical mechanisms of mind sometimes extend beyond the traditional boundaries of skull and skin.⁶ Although nothing in what I've just said strictly requires that the external elements in an extended cognitive system be technological artefacts (e.g. notebooks that provide external information-storage for extended memory systems, handheld computers that execute steps in extended reasoning processes), that is the standard way of the filling in the picture. Notice that in asking whether or not a particular

⁶ Clark and Chalmers (1998) provide the canonical statement and defence of ExC. For various recent developments and discussions of the view, see Menary 2010. For the dispute so far over whether enactivism is consistent with ExC, see, e.g., Clark 2008, Wheeler 2010, the contributions to Kiverstein and Clark 2009 (especially Di Paolo 2009 and Thompson and Stapleton 2009), Rowlands forthcoming, and Cappuccio and Wheeler forthcoming.

interpretation of deep continuity is consistent with ExC, I am not assuming that ExC is correct – and thus that any interpretation of deep continuity that is inconsistent with that hypothesis must be wrong – only that it is instructive, as a way of understanding the character, implications and relative merits of different interpretations of deep continuity, to determine which theoretical options those interpretations leave open and which they close off.⁷

The two enactive interpretations of deep continuity that I shall consider here endorse the following cluster of claims and inferences.

- Sense-making requires not only autonomy but also adaptivity; cognition is sense-making; so cognition requires not only autonomy but also adaptivity.
- Autopoiesis is sufficient for autonomy.
- Immanent teleology is necessary for life; autopoiesis is necessary for immanent teleology; so autopoiesis is necessary for life.

Assuming this shared theoretical backdrop, let's first revisit the position that, I have suggested, is Thompson's final view. This view is distinguished by the conjunction of two further claims – first, that life is autopoiesis plus cognition, and second, that autopoiesis is not necessary for autonomy. So, what is the nature of the enrichment process that defines the relevant sense of life-mind continuity? Put crudely, on the present view, life is what happens when autopoiesis is added to cognition, that is, when adaptive autonomy becomes autopoietic autonomy. And what that means is that, far from cognition being an enrichment of life, life is an enrichment of cognition. Assuming (as is surely natural) that cognition signals a phenomenon that we ought properly to call mental, the title of Thompson's book now takes on its full significance. All living systems are (to use an ugly term that Thompson doesn't) *minded*. So when Thompson deploys notions such as needful freedom, self-transcendence and immanent purposiveness to characterize life, and describes those existential

⁷ It is sometimes suggested that since, on the enactive view, cognition is relationally defined, it makes no sense to speak of cognition being spatially located at all (see e.g. di Paolo 2009; Thompson and Stapleton 2009). To put the point bluntly, cognition isn't extended because it isn't anywhere! This seems to me to be an almighty red herring. There is simply no problem about relationally defined phenomena exhibiting spatiality. The thing on which my laptop is sitting right now counts as a table only because of relational factors to do with the way it enters into a context of human activity, but I have no doubt at all that that table is located in space. Or, if the mention of an object makes you suspicious, the bodily movements I'm currently making count as typing only because of relational factors to do with the way they enter into a context of human activity, but I have no doubt at all that that typing is taking place in space.

structures as analogous to their psychological counterparts, the analogous character of the relation does not signal any loss of (to coin an even uglier term) *mindfulness*. The enrichment that takes place, such that “certain existential structures of human life are an enriched version of those constitutive of all life” (p.157), is an enrichment *within* the realm of the mental, and not an enrichment that, at some (perhaps indeterminate) point, *ushers in* the mental. That surely runs counter to expectations.

Is this view consistent with ExC? The strict answer, I think, is yes. Cognition is sense-making, which is adaptive autonomy, but since autopoiesis is not necessary for autonomy, and adaptivity is a modulation of autonomy, there could conceivably exist adaptively autonomous, that is cognitive, systems that are not themselves alive. Since, for Thompson, life is autopoiesis plus cognition, every living system is necessarily itself a cognitive system; but not every cognitive system is necessarily itself a living system. Thus there might exist an extended adaptively autonomous system – that is, *an extended cognitive system* – that although not itself alive, nevertheless had living elements, such as biological cells, among its components. So far so good. Unfortunately, however, if we reflect on the teleological structures that would characterize such a system, what we uncover is an unappealing – one might even say intolerable – conceptual mess.

To warm up our intuitions, let’s begin by imagining an adaptively autonomous (i.e., a sense-making) robot with no autopoietic elements. By the lights of Thompson’s account, this robot would qualify as an artificial cognitive system. However, if autopoiesis is necessary for immanent teleology, then the teleological structures applicable to this system must be extrinsic rather than immanent. Certainly nothing in Thompson’s account prevents there from being a cognitive system characterized by extrinsic teleology. But now let’s consider our extended cognitive system. For Thompson, it seems that this system, as a result of its hybrid organic-technological nature, must realize a kind of hybrid immanent-extrinsic teleology. But it’s genuinely difficult to see how to make sense of this idea. One avenue of thought might perhaps be that the extrinsic teleology enjoyed by the external technological elements is derived from the immanent teleology realized by the organic part of the system. Interestingly, this suggestion recalls a point sometimes made in relation to representational content by *opponents* of ExC. The critic argues that the contents carried by putatively cognitive extracranial elements are merely derived in character, and are thus parasitic on the nonderived content carried by neural states and processes. The implication (so the argument goes) is that the real cognition remains resolutely

internal. Adams and Aizawa, for example, argue that cognition always “involves non-derived representations, representations that mean what they do independently of other representational or intentional capacities” (Adams and Aizawa 2008, p.31). If the existence of hybrid representational content can be worked up into a point against ExC, then the same might be true of the kind of hybrid teleology with which we are currently concerned. And that seems to set up a potentially disagreeable clash with the thought that cognition as enactive sense-making (adaptive autonomy) is consistent with ExC.

It might seem that all is not lost. For at this point in the established content-centred debate, the advocate of ExC might well argue that, pace Adams and Aizawa, a genuinely cognitive trait may feature elements that carry only derived content, just so long as those elements are part of an integrated process in which there is a contribution from other elements (e.g. neural ones) that carry nonderived content. But even if this kind of ExC-friendly maneuver is plausible in relation to functionalist-based content realization in extended cognitive systems, the parallel move in relation to enactive teleology seems positively misconceived. After all, if the relevant teleological structures here are brought forth by the adaptive behaviour of the autonomous system as a whole – which is surely what the enactivist is disposed to say – the proposed strategy of partitioning, in which different kinds of teleology attach to different parts of the system, looks to be predicated on something akin to a category mistake. Of course, one might tidy up the whole mess here by adopting the view (scouted above) that autonomy is sufficient for immanent teleology, since then the extended adaptively autonomous cognitive system would realize immanent teleology and the problems I have highlighted would go away. But that ‘solution’ would come at a high price for Thompson, in that it would place even more pressure on the thought that autopoiesis plays an essential role in our understanding of life and mind.

In light of the difficulties faced by (what I take to be) Thompson’s own considered view of deep continuity, let’s recall the anomalous position on autopoiesis and autonomy that, I argued earlier, may be found in the pages of *Mind in Life*, and see if an account of deep continuity based on that position fares any better. The anomalous claim was that autopoiesis is necessary, as well as sufficient, for autonomy. If this is right, then since cognition is adaptive autonomy, autopoiesis is also necessary for cognition. In effect, this new dependency would mean that cognition is adaptive autopoiesis, thus rendering it rather peculiar to claim, as Thompson does, that life is autopoiesis *plus* cognition, since all bona fide cognitive systems would already be autopoietic. Indeed, since

it now seems that both life and cognition are a matter of adaptive autopoiesis, one might as well say that life is identical with cognition. This would mean (i) that there are no noncognitive living entities, and thus that any process of enrichment must once again take place *within* the realm of the mental, and (ii) that there are no cognitive nonliving entities, and thus, assuming that autopoiesis remains a property only of organic cells, that there are no hybrid organic-technological extended cognitive systems. The first conclusion flies in the face of the most natural understanding of life-mind continuity, while the second displays an inconsistency between enactivism and ExC.

I have claimed that the natural understanding of deep continuity is that there are certain noncognitive properties of living entities that, when enriched in specific ways, generate the phenomena of mind and cognition, phenomena that are exhibited only by a subset of living things. This is, as we have seen, an interpretation of the idea that is closed off to Thompson, who holds that life is, in a tangible sense, minded all the way down. One might wonder whether this tension ought to matter to Thompson. After all, our most natural ways of thinking are sometimes wrong. But it is at least arguable that the more natural understanding of deep continuity is also more plausible than Thompson's picture. Indeed, one cannot help thinking that, however much we are told that Thompson's deflationary notion of minimal cognition should not be thought to encompass higher-order intentional states of belief or consciousness, and however much we are warned to think of needful freedom and self-transcendence as 'no more than' analogues of their existential counterparts in conscious human experience, the fact remains that if the language of cognition and its existential structures is to have any real theoretical bite in the case of the basic phenomena of life, the result is a rather unhelpful dose of something akin to panpsychism that doesn't so much solve the problem of the genesis of mind as throw a cloak over the thought that there is a genuine problem to be solved. If this is right, then we would be well advised to look for an alternative theoretical framework for deep continuity that supports the natural reading. I shall end this paper with a brief sketch of just such an alternative.

5. A Broadly Representational Approach to Deep Continuity

First let's consider one way in which genes might be said to *code for* (i.e. *represent in an outcome-directed manner*) phenotypic traits during biological development (Wheeler and Clark 1999, 2008). The point of departure for this account of genetic coding is that the developmental construction of phenotypic form is often to be explained by the operation of massively distributed causal systems

encompassing genes, nongenetic organismic elements and/or wider environmental factors. A much-cited example is sex determination in the Mississippi alligator. These creatures lay their eggs in nests of rotting vegetation which generate heat in varying quantities. Eggs that develop at lower temperatures (within some overall range) end up producing females, whilst those that develop at higher temperatures end up producing males. Environmental temperature is thus a nongenetic factor, the inclusion of which in our explanation is necessary, if we are to account for the details of phenotypic form. Now, given this sort of distributed developmental solution, if one thought that what it means for a gene (or a complex of genes) to code for a phenotypic trait is for that gene (or complex of genes) to fully specify the form of that trait, where 'full specification' requires that the form of a trait may be predicted purely on the basis of what may be known about the genes in question, then the claim that genes code for traits during development would be false. In the sort of distributed developmental system with which we are concerned, knowing the entire sequence of an organism's DNA will not be sufficient to predict phenotypic form. Fortunately, this understanding of genetic coding is far from mandatory. After all, in familiar cases of algorithms, programs, instruction-sets, and other outcome-coding elements, those states and processes are able to perform their outcome-generating functions only given some assumed backdrop of other causally active states and processes (e.g. working operating systems) that themselves bear some of the responsibility for the exact form of the eventual product. Genes may play their coding role in a similar way, that is, by functioning within an assumed ecological backdrop of other developmentally crucial states and processes extended over genes, nongenetic organismic elements and/or wider environmental factors

Of course, in view of this spreading of the explanatory load, one needs to say precisely what it is that makes the genetic contribution a distinctively representational one. One proposal here (e.g. Wheeler and Clark 1999) is that there are architectural features exhibited by the mechanisms of protein synthesis that, when taken together, mandate a broadly representational picture of the genetic contribution to development. For example, the mappings from particular nucleotide triplets of mRNA (as derived from DNA by transcription) to particular amino acids during protein synthesis are famously thought to be *arbitrary*, in the sense (tantamount to a kind of *information-driven multiple realizability*) that the class of equivalent states and processes that could play the same *functional* role in the developmental process in question is fixed by *informational* rather than first-order-physical factors. In addition, the mechanisms of protein synthesis involve a *functionally distinct* (although physically

distributed) subsystemic *module* that implements the process of translation by consuming *information* from the upstream transcription module. The idea, then, is that where one finds an architecture with these sorts of features, a representational understanding of that architecture will be both appropriate and explanatorily powerful.

There is obviously much to be said about the justification for, and the details of, this picture.⁸ For the present, however, what matters is that *if* something like it is right, then we can glimpse deep continuity in its more plausible form. Consider the way in which neural states might be said to code for (i.e. represent in an outcome-directed manner) behaviour during intelligent perception and action (see e.g. Wheeler and Clark 1999, Wheeler 2005). In harmony with the hypothesis of extended cognition (see above), the systems that explain perceptually guided action-generation will often be massively distributed over brain, body and world. For example, as John Haugeland (1995/1998) once noted, one way to succeed in driving to San Jose would be to consult a cognitive map of the route, that is, to access a stored neural representation that specifies how to get there. An alternative method would be to select the correct road, and then follow the signs until you arrive. In the second case, the driver's innards and the road collaborate as partners in the successful completion of the activity. Indeed, it is not merely that the environment is a cheap and up-to-date source of information (although it is that), but that any adequate characterization of the action-generation mechanism at work here would plausibly need to count the contribution of the road as being similar to that of the neural representation cited in the first solution. Thus, as Haugeland (1995/1998, p.235) puts it, "much as an internal map or program, learned and stored in memory, would... have to be deemed *part* of an intelligent system that used it to get to San Jose, so... the *road* should be considered *integral* to [the] ability".

Where perceptually guided intelligent action displays the sort of distributed solution nicely depicted in Haugeland's example, the neural states and processes involved will not fully specify the behavioural outcome. Nevertheless, the representational credentials of their contribution (where such credentials are warranted) may well be secured by the same kinds of features as were operative in the case of biological development (again, see e.g. Wheeler and Clark 1999,

⁸ Indeed, I myself have argued that once the details of the proposed account are filled in, it turns out that, strictly speaking, it is not the molecules of DNA that code in development, but rather the downstream nucleotide triplets out of which molecules of mRNA are constructed (Wheeler 2006). Moreover, I have conceded that Godfrey-Smith (2000) may be right to think that there are problems in extending coding talk beyond proteins to phenotypic traits (Wheeler 2006).

Wheeler 2005). Thus neural factors may perform an outcome-directed (action-oriented) representational contribution within an assumed ecological backdrop of other behaviourally relevant states and processes extended over the brain, the body and/or the world. And that representational contribution may be illuminated precisely as such by architectural properties such as arbitrariness and information-consuming modularity. The former is in truth little more than the kind of information-driven multiple realizability that underwrites broadly functionalist theorizing in psychology, while the latter, stripped of too much Fodorian baggage (Fodor 1983), figures in a wide range of evolutionarily and developmentally sensitive regions of contemporary cognitive science, from mainstream evolutionary psychology to evolutionary robotics to developmental cognitive neuroscience and beyond (for discussion, see Wheeler and Clark 2008). Both are wholly consistent with ExC (again, see Wheeler and Clark 2008).

The picture I have just drawn is incomplete in all sorts of ways. It is, as I have said, no more than a sketch. Perhaps the most salient worry in the current context is that it exploits a set of interconnected concepts (representation, information, function, modularity etc.) of which Thompson is hugely suspicious and largely critical (see e.g. chapter 7 of *Mind in Life*). It is beyond the scope of the present treatment to develop any sort of response to Thompson's concerns. So my twofold point here remains merely conditional. First, if the broadly representationalist picture I have sketched is right, then the basic structure of ecologically embedded representation that characterizes the genetic contribution to development is replicated at the cognitive level of the neural contribution to online action-generation. In my book, this is an example of deep continuity. To borrow Thompson's words, it is a case in which life and mind share a set of basic organizational principles, and the organizational properties distinctive of mind are an enriched version of those fundamental to life. Secondly, if my broadly representationalist picture is right, then in line with the more plausible reading of deep continuity identified earlier, there is no reason to think of genes as in any way minded. The phenomenon of enrichment that the view encompasses involves a bottom-up process in which certain fundamental structures from nonminded life are evolutionarily and developmentally incorporated into the basic structures of minded life. In other words, according to the account of deep continuity I favour, enrichment takes us *from* life *to* mind, it does not operate *within* the domain of mind. This leads me to conclude that, at least in terms of such diachronic incorporation, life is in mind, but mind is not in life.

References

- Adams, F. and Aizawa, K. (2008) *The Bounds of Cognition*, Malden, MA and Oxford: Blackwell.
- Cappuccio, M. and Wheeler, M. (forthcoming) The sign of the hand: Symbolic practices and the extended mind, *Versus*.
- Clark, A. (2008) Pressing the flesh: A tension in the study of the embodied, embedded mind?, *Philosophy and Phenomenological Research*, 76 (1), pp.37-59.
- Clark, A. and Chalmers, D. (1998) The extended mind, *Analysis*, 58, pp.7-19
- Di Paolo, E.A. (2005) Autopoiesis, adaptivity, teleology, agency, *Phenomenology and the Cognitive Sciences* 4(4), pp.429-52.
- Di Paolo, E.A. (2009) Extended life, *Topoi*, 28, pp.9-21.
- Fodor, J. A. (1983) *The Modularity of Mind*, Cambridge, MA: MIT Press.
- Godfrey-Smith, P. (1994) Spencer and Dewey on life and mind, in Brooks, R. and Maes, P. (eds.), *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, Cambridge, MA: MIT Press.
- Godfrey-Smith, P. (2000) On the theoretical role of genetic coding, *Philosophy of Science*, 67: 26-44.
- Haugeland, J. (1995/1998) Mind embodied and embedded, in his *Having Thought: Essays in the Metaphysics of Mind*. Cambridge, MA: Harvard University Press.
- Heidegger, M. (1927) *Being and Time*, Trans. Macquarrie, J. and Robinson, E. (1962). Oxford: Blackwell.
- Jonas, H. (1966) *The Phenomenon of Life: Toward a Philosophical Biology*, Chicago: University of Chicago Press. Reprinted by Northwestern University Press. 2000.

- Kant, I. (1790) *The Critique of Judgement*, Trans. Meredith, J.C., Oxford: Oxford University Press (this edition first published in 1952).
- Kiverstein, J. & Clark, A. (eds.) (2009) *Topoi*, 28, special issue on The Enacted Mind and The Extended Mind.
- Maturana, H. R. & Varela, F. J., (1980) *Autopoiesis and Cognition: the Realization of the Living*, Dordrecht: Reidel, Boston Studies in the Philosophy of Science, Vol. 43, D.
- Menary, R. (ed.) (2010) *The Extended Mind*, Cambridge, MA: MIT Press.
- Rowlands, M. (forthcoming) *The New Science of the Mind: from Extended Mind to Embodied Phenomenology*, Cambridge, MA: MIT Press.
- Thompson, E. (2007) *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, Cambridge, MA: Harvard University Press.
- Thompson, E. and Stapleton, M. (2009) Making sense of sense-making: Reflections on enactive and extended mind theories, *Topoi*, 28, pp.23–30.
- Varela, F. J. (1979) *Principles of Biological Autonomy*, New York: Elsevier North Holland.
- Varela, F. J., (1991) Organism: a meshwork of selfless selves, in Tauber, A. (ed.) *Organism and the Origin of Self*, Dordrecht: Kluwer Academic Publishers.
- Varela, F. J., Thompson, E. and Rosch, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.
- Wheeler, M. (1997) Cognition's coming home: the reunion of life and mind. In Husbands, P. and Harvey, I. (eds.) *Proceedings of the 4th European Conference on Artificial Life*, Cambridge, MA: MIT Press.
- Wheeler, M. (2005) *Reconstructing the Cognitive World: the Next Step*, Cambridge, MA: MIT Press.
- Wheeler, M. (2010) 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. and Clark, A. (1999) Genic representation: reconciling content and causal complexity, *British Journal for the Philosophy of Science*, 50(1), 103–135.

Wheeler, M. and Clark, A. (2008) Culture, embodiment and genes: unravelling the triple helix, *Philosophical Transactions of the Royal Society series B*, 363, 3563–3575.