Evan Thompson

Reply to Commentaries

Let me express my deep thanks to the contributors for taking the time to read my book, *Mind in Life*, and for writing their thoughtful commentaries, from which I have learned a great deal. Special thanks are due to Tobias Schlicht, whose hard work and dedication made this volume possible. In what follows, I will respond singly to each contributor (in alphabetical order) and do my best to address their main points. My replies to the commentators will be longer or shorter depending on the points they raised. (Unless otherwise noted, all parenthetical page references are to *Mind in Life*.)

Daniel C. Dennett

I would like to begin my response to Dennett on an autobiographical note. In 1990–91 I spent a year as a postdoctoral research fellow at Dennett's Center for Cognitive Studies at Tufts University. At that time I was in the last stages of writing *The Embodied Mind* (Varela *et al.*, 1991), and I was also at work on several papers and a book on colour vision (Thompson *et al.*, 1992; Thompson, 1992; Thompson, 1995). Dennett was running a seminar based on the manuscript of his book, *Consciousness Explained*. I learned a huge amount from Dennett that year, not just from his seminar, but from his exceptionally generous mentoring, which included numerous helpful conversations and critical reactions to my writing, as well as many introductions to leading philosophers and scientists. Since that time and to this day, Dennett's writings have stood for me as a model of how to do philosophy in dialogue and collaboration with science, and his views on many issues have served as a critical foil for my own thinking.

Correspondence:
Evan Thompson, Department of Philosophy, University of Toronto, Toronto, ON M5R 2M8, Canada *Email: evan.thompson@utoronto.ca*

Journal of Consciousness Studies, **18**, No. 5–6, 2011, pp. ??–??

What Dennett and I share and where we differ can already be seen by comparing *The Embodied Mind* and *Consciousness Explained* (these two books were written around the same time, so we were not able to take account of each other). Both books tackled the problem of consciousness or lived experience from the vantage point of cognitive science, and both discussed many of the same topics — the nature of the self and personal identity, the temporality of conscious experience, the brain as a complex system, colour perception, and evolution, among others. But whereas my co-authors and I advocated phenomenology and first-person methods of examining experience as a necessary complement to cognitive science, Dennett promoted heterophenomenology; whereas we criticized adaptationism, Dennett defended it; and whereas we advanced an embodied approach, Dennett upheld functionalism.

In writing *Mind in Life*, I found myself again confronting these differences and wanting to work through them in light of recent work. Hence I gave special consideration to Dennett's views throughout *Mind in Life*, specifically when I discussed evolutionary theory, the mental imagery debate, and the place of phenomenology in cognitive science and the scientific investigation of consciousness. Although much of what I wrote was critical, the criticisms reflect how important and fruitful I find Dennett's work for my own efforts.

Mind in Life was written in the wake of the death of my other mentor and close family friend, Francisco Varela, who had guided me intellectually and personally since I was 15 years old. One way to mourn a loved one's passing and to try to keep their memory alive is to take up their causes and fight for them. Reading *Mind in Life* now — ten years after Varela's death and four years since the book's publication — I can see how that emotion and sense of purpose played a strong role in shaping the form and content of some of what I wrote. I refer specifically to my treatment of the issues that divided Varela and Dennett — autopoiesis versus selfish genes, the autonomy perspective versus reverse engineering, and neurophenomenology versus heterophenomenology. I suspect this driving emotion lies behind the 'rathering' that Dennett reads in some of what I wrote. Were I writing *Mind in Life* today, I would try to give a more nuanced and balanced presentation.

Nevertheless, when it comes to Dennett's commentary, I have to say that it seems to me to amount mostly to opinion and rhetoric rather than argument (there's a 'rathering' for you). Moreover, he seems to misunderstand many of my points, and he sometimes states what he

takes to be criticisms of my view whereas they are actually points I made myself and took some time and care to discuss.

In what follows, I aim to identify the rhetoric, point out the misunderstandings, and re-state the substantive disagreements, so the reader can judge for herself the relative merits of our positions.

Dennett's rhetorical strategy is to identify himself with 'orthodoxy', while styling himself a 'reform-minded critic', and to brand me as a 'radical' or 'revolutionary'. He first used this strategy years ago in his two book reviews of *The Embodied Mind* (Dennett, 1992; 1993). Although these labels might have made some sense in the early 1990s — and did admittedly fit my co-author Varela's sense of himself as a radical Chilean contesting the hegemony of Anglo orthodoxy — they no longer seem apt. Cognitive science, philosophy of mind, and theoretical biology have changed considerably over the past decades, so that the contemporary scene contains a plurality of theoretical perspectives instead of a monolithic 'orthodoxy' challenged by a few 'radicals' and 'revolutionaries'. For this reason, at the beginning of *Mind in Life*, I present the history of cognitive science from cognitivism to connectionism to embodied dynamicism as a diversification of perspectives and methods growing from the progressive recognition of the importance of context, lived experience, and temporal dynamics for understanding cognition. As I write: 'In contemporary research, all three approaches [cognitivism, connectionism, embodied dynamicism] co-exist, both separately and in various hybrid forms' (p. 4). Nowhere in *Mind in Life* do I set myself up as a 'revolutionary' battling 'orthodoxy'. These categories have little meaning for my project, however convenient and comfortable Dennett may find them. What does have meaning for my project is to give voice to certain scientific and philosophical traditions that are invaluable for understanding mind and life — traditions that have been neglected in certain quarters and that provide important correctives against a number of misguided tendencies in contemporary thought (e.g. genocentrism in evolutionary theory, what I call 'informational dualism' in biology and cognitive science, and the neglect of serious and careful phenomenology in the science of consciousness).

Dennett's rhetorical strategy misdirects the reader away from serious consideration of the issues. The 'revolutionaries', he says, 'are not really so revolutionary after all', and 'reform-minded critics — myself among them — have already pointed out the caveats that pre-empt these assaults on orthodoxy'. Here we see the familiar ploy of those who do not wish to confront seriously another way of thinking that challenges their own: acknowledge the rival viewpoint with

an air of *noblesse oblige*, circumscribe its force by claiming to have allowed for it already, congratulate yourself for being so undogmatic and fair-minded, and then go back to doing things exactly as you had been doing them before.

Let me respond now to the specific points Dennett makes about the four themes in my book he chose to discuss.

1. Autopoiesis

First, a clarification. Dennett misreads me when he writes that I propose 'autopoiesis as a radical new foundation for evolutionary theory'. Autopoiesis is *one* key element in the account of life I offer; others are evolutionary-developmental biology ('evo-devo'), developmental systems theory, and theories of biological self-organization (besides that of autopoiesis). I do not make autopoiesis foundational for these other theories; instead, I weave them together into an enactive perspective on evolution. At the beginning of Chapter Five, I distinguish three complementary approaches to characterizing life — the evolutionary, the ecological, and the individual (pp. 95–7). The concept of autopoiesis targets life at the individual level and aims to characterize the minimal organization necessary and sufficient for a system to be living (there are problems with the sufficiency claim, however, as I discuss on pp. 122–7). Later in the chapter (pp. 118–9), I state that life at the individual level always has to be seen as ecologically embedded; and, at the beginning of Chapter Seven (pp. 166–7), I make the point that the single, individual organism here and now is an abstraction from both the organism as an ecologically embedded life-cycle and from the organism as a member of a reproductive and evolutionary lineage. I follow with a section devoted to explaining the links between autopoiesis, reproduction, and heredity, in order to show how the characterization of life at the individual level relates to the characterization of life at the level of reproductive populations. In the rest of the chapter, I use both the theory of autopoiesis and developmental systems theory to criticize genocentrism, and I use ideas from theories of biological self-organization to criticize certain mischaracterizations of natural selection. Dennett's statement that I use the theory of autopoiesis 'as a radical new foundation for evolutionary theory' is thus an inaccurate simplification of what I write.

Dennett writes that the theory of autopoiesis is virtually synonymous with Tibor Ganti's chemoton theory; that both basically amount to a generalized and deepened version of the cell theory; that the theory of autopoiesis does not contain anything that would particularly

inform a mainstream cell biologist; and that 'it doesn't predict anything in biology that hadn't already been well understood by earlier theorists, or dissolve any puzzles that had been bedevilling those theorists'.

Dennett is right that there are similarities between Ganti's concept of the chemoton and Maturana and Varela's concept of autopoiesis; we could also add Robert Rosen's metabolism-repair (M,R) systems to the list. (In *Mind in Life*, I discuss the relationship between autopoiesis and M,R systems, but not the chemoton theory; Pier Luigi Luisi, 2010, however, has reviewed the relationship between autopoiesis and the chemoton.) But Dennett's dismissive assessment of autopoiesis (and, by extension, perhaps also the chemoton theory) misses the mark. The cell theory states that the cell is the fundamental structural and functional unit of living organisms, and that all cells arise from pre-existing cells. But this theory does not explain the organization that makes the cell an individual, how that self-producing organization could be instantiated as a network of processes; and how that kind of organization could in principle self-assemble from other simpler processes and structures. All three approaches — autopoiesis, M,R systems, and the chemoton — take this problem of the living organization as their explanatory target. As I explain in *Mind in Life*, Darwinian thinking, in its classical and modern molecular forms, missed this problem about organization (a problem going back to Kant's discussion of the organism as a 'natural purpose'). Here is a case where, contrary to Dennett, the problem was indeed 'unnoticed', not simply 'underestimated' (an assessment I justify in some detail in *Mind in Life*). Today the problem occupies a central place in research on the origins of life and synthetic biology (Luisi, 2010), as well as artificial life and computational biology (see the 2004 special issue of *Artificial Life* devoted to autopoiesis). In these research fields, the theory of autopoiesis has underpinned work on the creation of autocatalytic micelles as models of proto-cellular systems in the origins of life (Bachman *et al.*, 1992), and it has inspired and guided a significant body of work in artificial life (McMullin, 2004). More generally, the theory of autopoiesis has helped to shape theories and models, and make predictions, in research on the origins of life (Luisi, 2010), the chemical synthesis of minimal self-producing systems (Bachman *et al.*, 1992; Luisi, 2010), the computational simulation of self-producing systems (Bourgine and Stewart, 2004; McMullin, 2004), and the modelling of autonomous agency in natural and artificial systems (Barandiaran *et al.*, 2009). *Mind in Life* reviews much of this work. Does Dennett really think that all this hard and original

work was already well understood by earlier theorists and does not help with any of the puzzles that bedevilled them?

Dennett writes that there are exceptions to my claim that 'a cell stands out of a molecular soup by creating the boundaries that set it apart from what it is not', and he presents the following as examples: the boundary is semipermeable; which things count as inside and which things count as outside is not always clear; there are transition zones; when one autopoietic system enters another, it could be an invader or a symbiotic ally. Here Dennett presents something as if it were a criticism when it is actually a point I make myself. First, as I point out, 'boundary' is equivocal. It can refer to the material-spatial boundary of a membrane, or it can refer to the topological-functional boundary that is determined by the system's organization. As I write: 'taking "boundary" to mean only a unicellular semipermeable membrane or even a multicellular epidermal layer seems too restrictive (plants and insects do not have a skin). Rather, the crucial matter is that the system produce and regulate its own internal topology and functional boundary, not the particular physical structure that realizes this boundary' (p. 107). I discuss this point in relation to both the issue of whether multicellular organisms count as first-order autopoietic systems (or instead as autonomous systems dependent on the autopoiesis of their cellular constituents — see pp. 105–7) and the issue of whether Gaia (the Earth's ecosphere) counts as an autopoietic system (as Lynn Margulis claims — see pp. 120–1). Here it is also worth mentioning that I do not hold, contrary to Dennett's suggestion, that 'only autopoietic systems can be the proper (literal, underived, etc.) bearers of various biological predicates'. Social systems such as ant colonies, beehives, and primate bands are not autopoietic; furthermore, it is an open question whether multicellular organisms or Gaia qualify as genuine first-order autopoietic systems. Second, the examples Dennett gives are not exceptions to the point I make about the cellular membrane; they point instead to the need to consider that boundary as plastic and as a participant in what I later call (following Ezequiel Di Paolo) the system's 'adaptivity' (the system's ability to regulate itself in relation to its milieu — see p. 148). I develop this point at greater length in more recent writings on the relationship between the enactive approach and the extended mind theory in cognitive science (Thompson and Stapleton, 2009; see also Di Paolo, 2009).

Another place where Dennett neglects what I write is when he asks why we cannot see systems both as sources of their own activity, specifying their own domains of interactions, and as transducers or

functions for converting input instructions into output products. Of course, we can look at systems in both ways, as I say myself: 'What counts as the system in any given case, and hence whether it is autonomous or heteronomous, is context-dependent and interest-relative. For any system it is always possible to adopt a heteronomy or external-control perspective, and this can be useful for many purposes' (p. 50). Dennett also neglects the immediately following sentences:

> Nevertheless, this stance does not illuminate — and indeed can obscure — certain observable patterns of behavior, namely, patterns arising from the system's internal dynamics rather than external parameters. An organism dynamically produces and maintains its own organization as an invariant through change, and thereby also brings forth its own domain of interaction… A heteronomy perspective does not provide an adequate framework to investigate and understand this phenomenon; an autonomy perspective is needed.

Dennett gives no argument to counter this point, but it is the crucial point at issue.

Dennett objects to my statement, 'natural selection is not an external force but the differential propagation of developmental systems' (p. 202). (Here he also misleadingly frames the issue as 'the major question of whether autopoiesis provides a genuine alternative to standard neo-Darwinism' — a claim I do not make for autopoiesis on its own, as I have already explained.) Dennett wonders what the mistake about selection is and whether anybody has ever made it, but here again he neglects what I write and the literature I cite (see pp. 206–9). The idea that evolution is a 'field of forces' and that natural selection acts as an outside force on the units of selection — the standard neo-Darwinian view — was explained in detail years ago by Eliot Sober in *The Nature of Selection* (Sober, 1984). My point is not that this conception is mistaken but that it is limited because it neglects the self-organizing dynamics of the objects of selection; when these dynamics are taken into consideration, the objects of selection have to be seen also as generators of selection; in this way, the idea of selection is no longer adequately described as an external force (Weber and Depew, 1995; 1996, have explained this conceptual development in detail). Furthermore, although I was not able to discuss this controversy in *Mind in Life*, for the past few years one of the most interesting issues in the philosophy of biology has been whether natural selection should be interpreted statistically as a bias in the mathematical aggregation of births, deaths, and matings — the view propounded by my colleagues Mohan Matthen and Dennis Walsh — or whether it should be interpreted as a cause of evolution. The enactive view of selection I

present in *Mind in Life* goes well with the statistical interpretation of natural selection, because it sees natural selection not as a cause but as the differential retention of inherited variation (or in developmental systems theory talk, as the differential propagation of developmental systems), which itself has many causes. In any case, clearly these issues about selection are important and substantive ones, but Dennett engages none of them in his commentary; he just acts incredulous that anyone could see things differently from the way he sees them.

Dennett objects to my saying that Dawkins' concept of an evolutionary arms race is a questionable metaphor; in Dennett's view, this concept is 'one of the most predictively fruitful insights in evolutionary biology'. Here Dennett seems to miss my point. When I write, 'the notion of an evolutionary "arms race" is merely a questionable metaphor taken from the realm of human affairs and projected onto the interactions between certain species' and that 'It is an entirely observer-relative description' (p. 205), what I mean is that 'arms race' is a loaded way of characterizing what can be more neutrally described as positive feedback in certain coupled dynamical systems in biological evolution (e.g. co-evolving genetic regulatory networks). What is predictively fruitful is not this metaphor but the mathematical models of the dynamical interactions. The metaphor is questionable because it projects a complex, human sociopolitical phenomenon onto biological evolution; in this way, it confuses the interpretive framework we use to look at phenomena with the phenomena themselves.

Dennett also neglects the context of my remark. That context is a critical assessment of the adaptationist idea that living systems become better 'designed' or 'adapted' to their niches through evolution by natural selection (e.g. through cumulative selection and evolutionary 'arms races'). Thus, immediately after my critical remark about the 'arms race' metaphor, I make the point that evolutionary theory provides no general variable property of 'adaptedness'. Instead, it uses a variety of significantly different technical measures of 'fitness'. Thus, talking about living systems becoming better 'adapted' to their environments through evolutionary 'arms races' has no clear meaning.

Dennett says nothing in response to these points. Instead, he challenges me to show 'a single instance in which autopoiesis (or developmental systems theory) has predicted or explained biological effects on a similar scale'. This challenge is misplaced for several reasons. First, the concept of autopoiesis has helped to shape theories and models, and make predictions, in a variety of fields, as mentioned above. Second, developmental systems theory is not a theory in the sense of a

specific model that makes predictions to be tested against other models; it is a general theoretical perspective, with roots in multiple experimental traditions in developmental biology and developmental psychology. Finally, to make a (valid) *tu quoque* point: the 'arms race' metaphor is not itself predictive or explanatory, for the reasons just indicated.

Dennett finds unbelievable my charge that he confuses heuristics and explanatory frameworks with the phenomena themselves. Yet, as we have just seen, this is precisely what he does with the 'arms race' metaphor. In objecting to my charge, Dennett again seems to miss my point and ignore the context of what I write. I am well aware that he has 'gone to considerable lengths over the years' to show that 'design-without-a-designer' is not a contradiction, and that 'the design stance works exactly as well for organisms and their parts and behaviours as it does for artefacts'. But my assertion is not that the design stance cannot be legitimately used for organisms or that the design stance implies that organisms are the products of intentional design. My accusation is that the design stance (reverse engineering) misses something fundamental about living systems — their autonomy — and that it is illegitimate to use this stance to argue that organisms *really are* natural artefacts (heteronomous products of design without a designer). The context of my remark is a consideration of the differences between functionalist and structuralist traditions in the history of biology and their relation to the distinction between viewing living systems as heteronomous (the reverse engineering approach) and viewing them as autonomous. In the functionalist tradition, which includes Paley's natural theology, Darwin's theory, and Dawkins and Dennett today, the governing concept is that of design and the organism is likened to an artefact, either as the result of intentional design (natural theology) or the blind watchmaker (Dawkins). Kant's criticism of applying the concepts of 'design' and 'artefact' to life, as well as the criticisms made in the traditions of structuralism and Rational Morphology, are not adequately answered by showing how there can be design-without-a-designer (see pp. 129–40, 210–1). The crucial point is that the concept of design does not bring into focus the autonomous organization proper to living beings. In *Mind in Life*, I build on that point and argue (i) that reverse engineering is a heuristic or interpretive stance that treats organisms as heteronomous systems; therefore, it does not provide the right kind of generalizations to talk about biological autonomy; and (ii) to claim that organisms *are* natural artefacts (the heteronomous products of design-without-a-designer) because they can be *interpreted* from a reverse engineering perspective

conflates that perspective with the phenomena themselves. Dennett's reminder that his design stance works for organisms as well as artefacts just amounts to saying that organisms can be *interpreted* from the design stance — a point I do not dispute (see, e.g. p. 460, note 22 for what I call 'a nice case of reverse engineering'). What I do argue, to repeat, is that interpreting them this way fails to reveal their autonomy.

Dennett says I present 'no support at all' for the confusion I attribute to him (which is not the 'elementary confusion' he takes me to be attributing to him, as I just explained). But the support is there in the points I make in the surrounding context (as just detailed), to which Dennett makes no reply; moreover, his remark in his commentary about evolutionary 'arms races' displays the same confusion. In addition, I supply a supporting footnote (note 24, p. 461), in which I explain (i) that Dennett's term 'Design Space' already biases the discussion in favour of the functionalist or reverse engineering perspective, whereas the theoretically more appropriate and neutral term for the phenomena under consideration is 'Morphospace', and (ii) that Dennett's view that modern biology would be impossible without adaptationism mistakenly treats adaptationism as an *a priori* condition of biology instead of as a specific research programme. Dennett does not reply to these points either.

Finally, when Dennett characterizes Maturana's approach to autopoiesis as a reverse engineering approach, and says that my clarification of autopoiesis reveals the 'functionalistic rationales' of that theory, he neglects what I say on pages 144–5 about autopoiesis, namely, that this theory is a theory of *organization*, not *function*, and that, according to this theory, the notion of function has no explanatory value in characterizing a system's autonomous organization.

2. Developmental Systems Theory

Dennett focuses in this section on my critical remarks about genocentrism and what I call 'informational dualism'. The issues here concern the role of genes in evolution and development, and the conception of genes as carriers of information. Genocentrism or gene selectionism holds that genes are units of digital information; that as units of information genes have a special causal status in evolution and development compared to all other non-informational factors; that genes are 'replicators' that construct and control 'interactors' (phenotypes); and that genes are the primary units of selection in evolution. Developmental Systems Theory argues that this conception of

genes is misguided: genes are not units of information in any sense of 'information' that would not also apply to non-genetic factors; the replicator/interactor distinction is unhelpful (if a replicator is defined as anything that reliably replicates, then there are many other replicators besides genes; if a replicator is defined as an entity that replicates itself through its own causal power, then the only replicator is the reproducing organism or life cycle); and the unit of selection in evolution is the life cycle.

It is striking that Dennett in his reply provides no arguments that join this debate. Instead, he makes a series of complaints — that I present a caricature of genocentrism (despite my having accurately reported what he and Dawkins say); that I enlist the help of too many theorists (despite the importance of their work, which he caricatures as 'revolutionary' and offhandedly dismisses); and that I do not spell out my claims about the misapplication of the concept of information to genes (despite my detailed discussion of precisely this matter on pages 54–7 and 179–87).

On pages 186–7 I quote two passages from Dawkins and one from Dennett that express informational dualism. Dawkins' passages concern life and DNA; Dennett's concerns consciousness and the identity of the person through time. Each treats information as ontologically distinct from its contingent material expression, as pre-existing that expression, and as not affected by that expression. This way of thinking about information, I claim, reifies information and is structurally isomorphic to vitalism and mind–body dualism.

Dennett says he does not understand my claim. Am I following those 'sciencephobes' who use 'the epithet "dualism" to attack any science that uses the concept of information'? Clearly not; my book contains a huge amount of science (and who are these unnamed 'sciencephobes' anyway?) Instead of attacking sciences that use the concept of information, I use science to try to correct certain misconceptions about information (see pp. 57–8). Surely, Dennett writes, I am not 'claiming that the hardware/software dualism of *computer science* is dualism of any objectionable (e.g. Cartesian!) kind'? Yes and no. The hardware/software *distinction* of computer science does not necessarily imply informational dualism, but the computationalist and functionalist *dualism* of mind as informational software versus brain as hardware is tantamount to informational dualism. As I write in *Mind in Life*, 'Genocentrism and computationalism… run on the same conceptual fuel' (p. 174; see also pp. 185–6). So, yes, I do object to the hardware/software dualism of computationalism as well as to the misapplication of this dualism to genes. Note that the computationalist

version is exactly the dualism Dennett expresses in the passage I quote on pages 186–7: 'If what you are is the program that runs on your brain's computer… then you could in principle survive the death of your body as intact as a program can survive the destruction of the computer on which it was created and first run' (Dennett, 1991, p. 430). Dennett seems blind to the fact that his dualism is indeed a dualism of a Cartesian kind — a surprising oversight from someone who coined the term 'Cartesian materialism' and who so trenchantly showed how neuroscience and psychology can remain objectionably Cartesian despite their being materialistic (*ibid.*). Of course, the computationalist version is not a substance dualism, but it remains markedly Cartesian in its disembodied conception of mental processes (Descartes' own view was actually more complex, as I discuss in Chapter Eight of *Mind in Life*).

Dennett says I never spell out my claim that the hardware/software dualism of computationalism is misapplied to genes. This charge is unjust. On pages 180–2 I explain the claim and give two supporting reasons with citations to the relevant literature (see also pp. 54–7). First, hardware and software are independent of each other in a way that DNA/RNA and the rest of the cell are not (the latter produce and depend on each other autopoietically; the former do not); and second, the notion of genes containing 'information' in the form of coded instructions for developmental outcomes is faulty because it confuses the causal specificity of the relations between DNA/RNA and amino acids with coding for phenotypic design (see pp. 57 and 181–2).

Dennett makes no reply to these points. Instead, he offers what seems to him 'a very clear and unobjectionable way in which we can draw the software/hardware distinction when discussing genes'. But all he mentions are the familiar facts that DNA triplets specify amino acids and that proteins are produced through the complex orchestration of RNA and ribosomes. For the reasons I explain in *Mind in Life*, these molecular relations of causal specificity do not warrant a conception of DNA sequences as software 'coding for' phenotypic characteristics as hardware (see pp. 54–7, 180–2). Dennett, however, says nothing about these reasons.

Dennett wonders whether anybody has ever subscribed to the myth of the gene as a unit of pure information. He seems to have forgotten Dawkins's words:

> After Watson and Crick, we know that genes themselves, within their minute internal structure, are long strings of pure digital information. (Dawkins, 1995, p. 17)

> Genes are pure information — information that can be encoded, recoded, and decoded, without any degradation or change of meaning… We — and that means all living things — are survival machines programmed to propagate the digital database that did the programming. Darwinism is now seen to be the survival of the survivors at the level of pure, digital code. (Dawkins, 1995, p. 19)

It is exactly this way of thinking that exemplifies the myth of the gene as a unit of pure information.

Dennett complains that although I cite and criticize a passage in which he discusses how genes do not carry information intrinsically, I do not address his claim that it is nonetheless possible to make a principled distinction between explicit (coded) and implicit (uncoded) information. But Dennett does not answer the question that I raise immediately after quoting him: 'If information from the environment is needed to make the genetic information informational in the first place, then what is the ground for holding onto the genocentric tenet that genes are the informational prime-movers?' (p. 184). I then quote a passage from Susan Oyama that details the problems with Dennett's conception of information; Dennett rejects the passage as 'a series of non-sequiturs' but neglects to tell us what those non-sequiturs are.

In defence of the genocentric claim that genes (at least sometimes) are informational prime-movers and cells their vehicles, Dennett offers the familiar fact that tiny changes in genes can yield huge downstream effects. But changes to non-genetic factors (such as methylation patterns) can also have large downstream consequences (see p. 177). Here too Dennett says nothing about the reasons and evidence I present against the replicator/interactor conception of genes and the cells that house them (see pp. 177–8, 197).

Dennett characterizes my marshalling of evidence and argument from a number of important biologists and philosophers as 'throwing the kitchen sink at orthodox neo-Darwinism', and he says that if this is the best I can come up with, then neo-Darwinism must be in pretty good shape. But two can play at that kind of rhetorical game: 'If casually dismissing theorists, name-calling and caricature ("science-phobes", "radical biologists", "a veritable Hall of Fame of would-be revolutionaries of biology"), stating opinions as if they were arguments, and not responding to evidence and reason are the best Dennett can come up with, then my case must be pretty solid after all.'

3. Autonomous Meaning-Construction

The crucial issue here is what it takes for a system to qualify as having genuine agency and a meaningful perspective on the world.

I agree with Dennett that AI systems can in principle be designed to 'muck about in the world and devise their own categories'. My claim, however, or rather the fundamental hypothesis of the enactive approach, is that such systems must be designed as autonomous agents in the full-blooded and technical sense of 'autonomy' I specify in *Mind in Life* (see pp. 44–6), and whose minimal paradigm is the autopoietic cell. Since the writing of *Mind in Life*, this sense of autonomy has been defined more precisely with application to AI systems by Froese and Ziemke (2009) and Barandiaran *et al.* (2009). As their framework makes clear, and as I argue in *Mind in Life* (p. 160), replicating molecules do not qualify as autonomous, and hence do not qualify as having agency.

Dennett in his reply, however, claims that the autonomy perspective 'is not required for the sense of sense-making', and he reasserts his claim that macromolecules are agents. Yet he provides no reasons to back up this claim and to counter the reasons I give in *Mind in Life* for rejecting it (see pp. 160–1). Instead, he gives two examples: 'Think of motor proteins — little porters trudging along on their actin or tubulin highways carrying freight to where it is needed. Think of proof-reading enzymes.' Yet neither example counts as a case of agency. Agency requires more than causally orchestrated behaviour — unless, that is, we are willing to allow the word 'agency' to become so vague that the planets going around the sun or sugar molecules crystallizing count as agents. According to the enactive approach, agency requires at a minimum that the system 'manage the flow of matter and energy through it so that it can, at the same time, regulate, modify, and control (i) internal self-constructive processes and (ii) processes of exchange with the environment' (Ruiz-Mirazo and Moreno, 2004, p. 240). Put another way, agency requires at a minimum that the system meet the following three conditions: (i) be an individual, in the sense of producing and maintaining its own organization ('individuality'); (ii) be the active source of its interactions, in the sense of modulating the parameters of its coupling with the environment on the basis of its internal (self-organized) activity ('interactional asymmetry'); and (iii) generate the norms for those interactions on the basis of its activity ('normativity') (see Barandiaran *et al.*, 2009). Replicating macromolecules such as DNA/RNA, motor proteins, and proof-reading enzymes do not meet these three criteria, so they do not qualify as agents in this rigorous and naturalistic sense. By contrast, chemotaxic bacteria do qualify as agents. One of the basic errors of selfish gene theory — which Dennett perpetuates — is that it metaphorically

projects a faulty conception of agency onto entities that do not meet the minimal criteria for being agents.

4. Phenomenology and Heterophenomenology

The personal/subpersonal distinction and self-reports about experience. Dennett says he agrees with me that when one makes a self-report about one's experience, 'One is describing one's subjectivity at the personal level in a way that is completely noncommittal about the subpersonal workings of one's brain' (p. 305). Nevertheless, he also thinks that subjects make 'unwitting reference' to entities whose status as real or fictional is to be decided by looking at what is going on in their brains. Here we disagree. If the self-reports are properly descriptive in form (rather than conjectural about the underlying causes of behaviour) — in other words, if they are proper phenomenological reports — then we should not interpret them as making this kind of unwitting reference. For example, Dennett maintains that heterophenomenology gives us 'the leverage to discover that… since there are no images being processed in the brain when subjects say they are rotating mental images, their heterophenomenological reports must be interpreted as unwitting fictions of a sort'. I maintain, however, that when subjects say they are rotating mental images, they are giving a personal-level description of their subjective experience of visualizing how things look; they are not expressing beliefs or making unwitting reference to what is going in their brain considered as a cognitive system. Despite his claim to the contrary, Dennett's way of interpreting the subjects does amount to forgetting or abandoning the personal/subpersonal distinction he introduced, for two reasons. First, he evaluates purely *descriptive* self-reports about the phenomenal character of experience on the basis of what is going on *causally* and subpersonally in the brain; and second, he evaluates the reports by looking to see whether the reported phenomenal *content* matches or fails to match the representational *format* of the neural processes. To put the point another way, my claim is that even if there are no pictorial representations in the brain, it does not follow that the intentional content of imagery experience is fictional in Dennett's sense. As far as I can see, Dennett does not counter the argument I present for this claim on pages 304–5; he just re-states the view my argument challenges.

Intentional acts and intentional objects. Here Dennett reads 'intentional' as meaning deliberate, instead of being directed toward an object. According to phenomenology, conscious experience is intentional in the

sense that it is about or directed toward an object, and every intentional object implies a correlative mental act that intends that object. The object as perceived implies a certain act of perception that perceives it that way, the object as remembered implies a certain act of remembering that remembers it that way, the object as imagined implies a certain of act of imagination that imagines it that way, and so on.

My point in *Mind in Life* (pp. 305–6) was twofold. First, we need to attend to both correlative sides of an intentional experience when we analyse that experience phenomenologically. In the case of mental imagery, for example, we need to characterize the intentional object — the shape as visualized — and we need to characterize the intentional act — the activity of visualizing (see pp. 291–7 for this analysis, and pp. 297–303 for its application to the mental imagery debate). Second, intentional acts such as perceiving, imagining, and remembering should not be identified with believing, which is another type of intentional act. Believing is what happens when subjects have to make a judgment. Thus, in the case of subjective experience, a belief about experience is what gets expressed when subjects are asked to make and report a judgment about their experience.

First-person methods. I do not deny that heterophenomenology can acknowledge the unavoidable need to make use of first-personal modes of access to mental phenomena. As I write (p. 306): 'There seems to be nothing in the heterophenomenological method that disallows using the first-person perspective in this direct phenomenological way… if the material on which heterophenomenology goes to work is first-person reports about experience, and if the production of such reports sometimes requires that subjects attend to and describe their experience, then heterophenomenology already depends on the first-person mode of access to mental phenomena being put to work in an experimental setting.' My critical point is that heterophenomenology has had virtually nothing to say about the idea — a keystone of phenomenology — that conscious experience needs to be explored from within the first-person perspective (either descriptively, as in phenomenological psychology, or transcendentally, as in transcendental phenomenology). Moreover, it has had nothing to say about the proposal to use first-person methods of training attention and awareness in order to sensitize individuals to their experience in ways enabling them to describe it more precisely (compared to individuals without that kind of mental training). This proposal is a key feature of neurophenomenology, in which one of the working hypotheses is that reports from individuals with such mental skills are useful for identifying the variability in the brain response on a trial by trial

basis, so that more precise correlations can be established between the spontaneous flow of consciousness and spatiotemporal patterns of intrinsic brain activity (Lutz and Thompson, 2003; Cosmelli *et al.*, 2007; Lutz *et al.*, 2008a). I like Dennett's liberal conception of heterophenomenology as giving 'the subject the best possible way to let it all hang out', but this allowance seems like lip-service given heterophenomenology's lack of consideration to first-person methods (except to deride their use in 'lone-wolf phenomenology', which no proper phenomenologist has ever advocated or practised). Heterophenomenology on its own does not provide the first-person phenomenological procedures needed for a proper examination of subjective experience; it presupposes them. For this reason, as I say in *Mind in Life* (p. 307), heterophenomenology by itself must be deemed methodologically incomplete.

With regard to the Lutz *et al.* (2002) study, which I prefer to think of as a pilot study and not a 'flagship attempt', I agree that it has the limitations that Piccinini (2010) mentions. Meanwhile, since the Lutz *et al.* pilot study, other neurophenomenological studies using first-person and second-person methods for probing experience have begun to appear (e.g. Christoff *et al.*, 2009; Farb *et al.*, 2007; Lutz *et al.*, 2008b; Petitmengin *et al.*, 2007), so this kind of research, although fledgling, looks promising.

Experience versus beliefs about experience. Here the disagreement is easy to state. My view is that properly descriptive statements about experience should be interpreted as direct expressions or verbalizations of those experiences and not as beliefs about those experiences. Interestingly, Piccinini (2010) proposes precisely this way of interpreting first-person reports as one of a number of changes for improving heterophenomenology. In his words:

> What remains unclear [in Dennett's procedures for doing heterophenomenology] is why we should assume that reports always express beliefs about the target mental states rather than expressing the target mental states themselves… A more prudent treatment of reports is at hand: establish to the extent possible that a report expresses the target mental state (e.g., an emotion, desire, memory, or what have you, including a belief if that's the target), then interpret the report as providing information about the target mental state. After all, that is what we are trying to investigate. (Piccinini, 2010)

I would add only that one point of first-person phenomenological methods is to enable subjects to express maximally and accurately their target mental states, instead of having them make judgments and thereby express their beliefs about their target mental states.

For some examples of first-person/second-person methods being used to gain access to implicit aspects of emotion and cognition, see Farb *et al.* (2007); Nielsen and Kaszniak (2006); and Petitmengin *et al.* (2007).

What's in a name (phenomenology or heterophenomenology)? Dennett says my own proposed method 'is really heterophenomenology after all *rather* than a radical alternative', and that if I will adopt his method, then he will adopt my name for it — second-person heterophenomenology. Let's set aside the epithet 'radical' as a distraction from the main issue. If Dennett's method now (i) permits using first-person methods for examining experience; (ii) encourages their use in certain contexts (e.g. for examining attention and emotion regulation in relation to neural plasticity, or for relating spontaneous fluctuations in subjective experience on a fine time-scale to neural measures on a trial by trial basis); (iii) treats phenomenological self-reports as verbalizations of target mental states rather than expressions of belief about those mental states (unless there is good reason to interpret them otherwise); and (iv) does not evaluate the truth or falsity of the content of such reports by examining whether they match or fail to match the representational format of the subpersonal neural processes, then there is no longer any substantive difference between the methods we propose. As for the name, 'phenomenology' seems to me less cumbersome and misleading than 'heterophenomenology', for as I write in *Mind in Life* (p. 307): 'Phenomenology from its start has already encompassed heterophenomenology (or its possibility).'

To close my response to Dennett, I think I have shown that *Mind in Life* is hardly about setting up strawmen or poking certain authors with a parody of their views. As for getting Dennett to dance the tango, I wish I could make it happen, because I think we all can benefit from learning some new moves.

Lucia Foglia and Rick Grush

In their commentary, Foglia and Grush ably defend an emulation account of mental imagery against a simulation account, and they ask whether I see the emulation account as a friendly clarification of my enactive account or as a challenge to it. The question arises because enactive accounts criticize representationalist theories of perception, and the emulation theory appeals to representations, understood as mental models. The short answer to their question is that I welcome their account and see it as a friendly supplement to my remarks about

sensorimotor processes in mental imagery (see pp. 295–6, 298–9). But let me say a little bit more.

First, the concept of representation is used in numerous ways in cognitive science and the philosophy of mind, as **Hutto** helpfully details in his commentary. I argue against representationalist theories that separate perception and action, instead of recognizing their constitutive interdependence, and that neglect the ways autonomous agents bring forth or enact meaning in perception and action (see pp. 10, 58–9). Since the emulation theory does not require these typical features of representationalism, my objections to representationalism need not apply to the emulation theory.

Second, appealing to emulation models to explain mental imagery still leaves open exactly how those models are realized in the brain. With regard to this issue, I have doubts about the way Grush (2004) applies control theory to the brain. Specifically, I doubt that the engineering distinction between 'controller' and 'plant' (controlled system) applies to the brain considered as a complex dynamical system (see Kelso and Kay, 1987; Kelso, 1995; and Appendix B to *Mind in Life*). This issue, however, goes well beyond Foglia and Grush's commentary, so I will not pursue it here.

Finally, Foglia and Grush's argument that mental imagery requires an internal model is reminiscent of Sartre's (2004) idea that imagery requires an 'analagon' that functions to direct the image consciousness to its intentional object. In perceiving a picture, the analagon is overt; in visualizing an object it belongs covertly to the image consciousness (though not as an object inside consciousness). This interesting correspondence between phenomenology and the emulation theory, regarding the role of 'models' in imagination, would be worth pursuing further.

Dan Hutto

Hutto aims to open up a dialogue between analytical philosophy of mind and the enactive approach. His strategy is to show how teleosemantic theories of content need to be modified in a variety of ways that end up bringing these theories closely in line with the basic orientation of the enactive approach. I welcome this dialogue, greatly appreciate Hutto's bridge-building efforts, and find myself largely in agreement with his commentary.

My one caveat concerns something Hutto himself mentions — the interpretation of evolution. Teleosemantic theories have traditionally worked with a strongly adaptationist view of evolution, in which

natural selection is conceived as an outside force acting on agents conceived in heteronomous terms (see my response to **Dennett**). In contrast, the enactive approach calls attention to autonomous systems as generators of selection, and to the need to understand selection as a kind of dynamical stabilization resulting from many causes (see also **Oyama's** commentary). Hence, Hutto's version of a modified teleosemantics — 'teleosemiotics' as he calls it — would also need to move away from adaptationist views of evolution in order to find common ground with the enactive approach. I see no reason why this movement cannot happen, though I suspect the resulting teleosemiotic theories would look rather far removed from their teleosemantic ancestors.

Albert Newen

Newen thinks my 'aims are much too ambitious' and he criticizes the path I chart from autopoiesis and sense-making to cognition and consciousness.

1. Autopoiesis, Adaptivity, and Cognition

One of the core ideas of *Mind in Life* is what I call the 'deep continuity of life and mind'. There are different ways to express this idea. One way is to say that to be a living system is also to be a sense-making system, and thus that life is sufficient for mind. More precisely stated, being an adaptive autopoietic system is necessary and sufficient for being a living system; and being an adaptive autopoietic system also suffices for being a sense-making system. To be a sense-making system is to be a cognitive system, in a wide or broad sense of the term 'cognitive'.

Newen challenges this thesis; in his view, adaptivity is the crucial property for cognition, not autopoiesis. He seems to misunderstand these concepts, however, including their relation to each other.

As the models of autocatalytic micelles and autopoietic tessellation automata show, *minimal autopoiesis* (a semipermeable boundary autocatalytically constructed by reactions remaining within the boundary but also occurring at the boundary) is sufficient for *individuality*, but not for *behaviour* in the sense of the system's modulating the boundary conditions and parameters of its coupling with the environment on the basis of its internal activity. In more concrete biological terms, minimal autopoiesis is necessary but not sufficient for such basic phenomena of life as metabolic assimilation and accomodation, or chemotaxis. Such phenomena also require what Ezequiel Di Paolo

(2005) calls 'adaptivity', namely, being able to modulate the autopoietic process in relation to conditions registered as improving or deteriorating, viable or unviable, for the system. As it happens, every naturally occurring autopoietic system we know is adaptive in this sense; nevertheless, as the above models show, it is possible to construct minimally autopoietic systems that are not adaptive.

The crucial point Newen misses, however, is that, in the case of an adaptive autopoietic system, adaptivity is not an external add-on to autopoiesis but a complexification of autopoiesis. In other words, autopoiesis is not a mere background condition; autopoiesis is constitutive of adaptivity. To state the point more generally, in the case of an adaptive autonomous system (the paradigm of which is an adaptive autopoietic system), autonomy is constitutive of adaptivity. For this reason, it is not the case that the explanations for the sense-making or cognitive abilities of living systems rely only on the property of being an adaptive system and not also on the property of being an autonomous system.

The first section of Newen's commentary contains a few other problems:

(1) Newen's declaration that bacteria, amoebae, and plants are not cognitive systems is question begging, for I maintain that these organisms are sense-making systems, and thus cognitive in a broad (but well motivated) sense of the term.

(2) My reference to 'internal self-production of the minimal sort' on page 129 is a reference to minimal autopoiesis; contrary to what Newen seems to imply when he quotes me, living cells are not merely minimally autopoietic but also adaptively autopoietic. Furthermore, bacteria, the simplest naturally occurring autopoietic systems, are highly adaptive, so it is not the case that such cells can survive only as long as the environment is very stable.

(3) It is misleading to say that I accept Bourgine and Stewart's (2004) tessellation automaton 'as a case of an autopoietic system that is neither a biological nor a cognitive system'. The tessellation automaton is a mathematical model of an autopoietic system, not an actual autopoietic system.

2. Does Autopoiesis Plus Adaptivity Entail Sense-Making?

Newen argues against my claim that autopoiesis plus adaptivity suffices for sense-making or what I call cognition in its minimal biological form. For Newen, cognition requires that the system be representational, specifically that it operates on the basis of internal representations of goal

states. Without such internal representations, he maintains, we have no norms but only mechanisms.

First, let me point out an apparent misreading. I do not call sense-making a 'narrow sense of cognition'. Rather, I say that living is a process of sense-making, of bringing forth significance and value, and I describe sense-making as cognition in a *broad* or *wide* sense of the term (p. 159).

Second, Newen's view has the following problems: (1) he helps himself to the notion of representation, as if it were obvious what this notion means, and without addressing my reasons for rejecting it, at least as this notion has been standardly understood (see **Hutto's** commentary for further discussion). From my point of view, his appeal to representations as the mark of the cognitive is question begging; (2) he presupposes that there is a meaningful way of contrasting mechanisms and norms. Yet one of the main points of my discussion of autonomous systems is that such systems are *organized* in such a way as to be normative (see also Barandiaran *et al.*, 2009). So the dichotomy between mechanistic reactions and norms does not apply to these kinds of systems in the way Newen assumes; (3) such normativity is not something inside the system (e.g. as the content of a representation of a goal); it is a relational property of the system and its environment.

In sum, Newen's criticisms in this section depend on assuming or presupposing concepts and distinctions I do not accept and that I subject to criticism in *Mind in Life*. Newen does not address those criticisms.

3. The Deep Continuity of Mind and Life

Newen thinks my deep continuity claim can be read in two ways, one weak and trivial, the other strong and false. Neither reading gets at the heart of the deep continuity thesis.

According to the weak reading, the organizational properties of mind are an enriched version of those fundamental to life. Newen does not make clear why this claim is trivial, but he does say that it is neither original nor controversial. I make no claim to originality. As for being controversial, if 'organizational properties' refers (as it does here) to self-organization, collective dynamics, circular causal processes, autopoiesis, and so on, then the claim is certainly controversial from the standpoint of cognitivist cognitive science as well as most versions of functionalism (see Clark, 2001, p. 118, quoted on pp. 128–9).

According to the strong reading, 'the principles of mind are essentially constituted by the principles of life'. I do not make this claim in *Mind in Life*. What I claim is (i) that the organizational principles of mind are an enriched version of those fundamental to life, and (ii) that there is an existential-phenomenological continuity of mind and life (p. 129). This second claim is the one I develop through a reading of Kant and Jonas.

But what about the question of whether the principles of mind are essentially constituted by the principles of life? Newen seems to think this claim is obviously false for two reasons. First, cognitive abilities require representations of goal states. I have already indicated that I reject this assumption (for reasons indicated in *Mind in Life*). Second, autopoietic principles are mere 'background conditions' for cognition but not essential to cognition. On this point, however, much more cautious consideration is needed (see also the commentaries by **Protevi**, **Wheeler**, and my replies).

I maintain that autopoiesis plus adaptivity is (i) necessary and sufficient for life, and (ii) sufficient for mind. It follows that mind is necessary for life. The question remains: is life necessary for mind?

On the one hand, I maintain that mind — sense-making, cognition, and consciousness — requires that the system be *autonomous* (in the precise sense of autonomy detailed in *Mind in Life*). I also allow that there can be autonomous systems that are not themselves autopoietic systems. So far, however, every autonomous system we know depends constitutively on autopoietic constituents.

But what about AI systems and robots? AI and robotics have not yet managed to create autonomous agents in the relevant sense of 'autonomy' (see Froese and Ziemke, 2009; Barandiaran *et al.*, 2009). Hence I disagree with Newen when he writes, 'Modern robots with complex behavioural abilities, e.g. in spatial navigation, have knowledge about their environment but they are so far not autopoietic systems'. Because these robots are not autonomous agents, they do not embody knowledge about their environments; the 'knowledge' is attributed and in the eye of the beholder (see Froese and Ziemke, 2009).

The crucial remaining question is whether autopoiesis is a necessary constitutive feature of autonomy; in other words, can there be autonomous systems that do not depend constitutively on autopoiesis? Put another way, does autonomy require metabolic self-construction?

Although some might think the answer to this question is obviously *no*, reasons can be given for thinking the answer might rather be *yes*. For example, perhaps sense-making requires the kind of selfhood and

concerned perspective on the world that comes from having constantly to renew oneself metabolically in precarious thermodynamic conditions. If there were this requirement, then autopoiesis would be necessary for autonomy, and so life would be necessary for mind. I postpone further discussion of this issue for my reply to **Wheeler**.

Were autopoiesis to turn out to be necessary for autonomy and hence for cognition, it would not follow that appealing to autopoiesis or to being a living system would suffice for explaining cognition. Nothing I write in *Mind in Life* implies that I maintain the implausible thesis that the principles of being a living a system suffice for explaining all of cognition, so it is not clear to me why Newen seems to think I believe this thesis.

4. Artificial Life

Here I need to correct two misunderstandings.

First, as already indicated, I do not maintain that the tessellation automaton is an autopoietic system that is not a biological system; I maintain that it is a mathematical model of minimal autopoiesis. I also call attention to the question of whether autopoiesis can be completely modelled this way when I discuss the relationship between autopoiesis and Robert Rosen's metabolism-repair systems (p. 144).

Second, it does not make sense to say that 'autopoietic organization is a structure'. An organization is a set of relations among processes; a structure is a concrete instantiation of those relations. The autopoietic organization can be concretely realized in a variety of structures, and any autopoietic system is constantly changing its structure (through the metabolic turnover of its constituents).

5. Can Life Be Known Only by Life?

Jonas's statement 'life can be known only by life' is a transcendental statement in the following sense: it is about the conditions for the possibility of knowing life, given that we do actually have biological knowledge (p. 164). Newen's claim that I do not distinguish clearly enough between knowledge in general and empathic knowledge misses the point. My claim is that in order to recognize or bring into focus a certain *form* or *pattern* of phenomena — the pattern of a self-producing unity relating flexibly and adaptively to its environment — we need to embody that pattern ourselves, i.e. to be living beings. I will not repeat Jonas's argument (or my version of his argument) for this claim here. Instead, I will indicate why I think Newen's considerations against transcendental thinking have no merit.

First, the transcendental framework I am working with in *Mind in Life* is not strictly Kantian; it belongs to transcendental phenomenology following Husserl and Merleau-Ponty. I explain this conception of the transcendental in Chapter Two. Newen's three considerations against transcendental philosophy, however, target Kant and do not speak to transcendental phenomenology: (i) transcendental phenomenology rejects the postulation of an unknowable thing-in-itself; (ii) transcendental phenomenology does not separate the transcendental ego and the empirical ego, but maintains that the transcendental ego is the empirical ego considered as a condition of possibility for intentional consciousness (see Sokolowski, 2000); (iii) transcendental phenomenology does not preclude an understanding of mental causation, but views mental causation in terms of the motivational relations between embodied intentional experiences.

Second, Newen's statement that 'modern philosophy of mind in its majority' has left behind transcendental thinking is simply false and hugely distorts philosophy of mind. Transcendental considerations abound in philosophers such as Wittgenstein, Davidson, Putnam, Evans, and McDowell, to name just a few. Furthermore, one of the most interesting and fertile philosophical developments in recent years combines transcendental and existential phenomenology with analytical philosophy of mind and philosophical psychology in order to make headway in understanding phenomenal consciousness (e.g. Zahavi, 2005).

6. Consciousness

Newen misreads my discussion of phenomenal consciousness and attributes to me claims I do not make. I never claim that autopoiesis is 'essential' for phenomenal consciousness. Instead, I show how certain existential structures of embodied experience have their roots in basic structures of biological life.

Newen writes that although human feelings are 'connected' with self-regulation, 'there is no evidence that this is essentially so'. It is not clear to me what he means by 'essentially'. Does he mean that it is conceptually or metaphysically possible for human emotions to occur in the absence of how the human body affectively regulates itself in relation to environmental events? What would be the argument for this highly implausible claim?

Similarly, Newen claims that 'sensorimotor couplings' are not essentially connected with phenomenal consciousness. Here I suggest

he read **Siewert's** commentary, which ably elaborates my view that bodily movement is constitutive of perceptual consciousness.

Newen states 'humans are already born with basic phenomenal consciousness that is active before intersubjective interaction can develop'. This statement overlooks the different aspects of intersubjectivity I discussed in Chapter Thirteen. The statement is true only if 'intersubjective interaction' refers to social cognition involving perspective-taking and mutual self/other understanding (pp. 395–401); it is false if 'intersubjective interaction' includes the 'primary intersubjectivity' of affective resonance and sensorimotor coupling (pp. 393–5; see also Gallagher and Meltzoff, 1996).

Newen also misreads my criticism of the zombie argument. First, I do not grant the logical possibility of zombies. It is not obvious to me at all that zombies are logically possible; to determine whether they are would require that we fill in all the important details in the zombie scenario, something proponents of the zombie argument never do. I strongly doubt, for the reasons given in *Mind in Life*, that the zombie scenario is conceivable (though I do not claim to have shown that it is inconceivable). Second, I do not criticize the zombie argument in order to show 'the non-reducibility of consciousness'. I criticize the zombie argument in order to reveal the highly problematic conception of phenomenal consciousness it presupposes.

Newen also misunderstands my claim that 'Consciousness, considered as epistemic base, is equivalent to the experiential acts by which… objects are disclosed to us' (p. 239). This is a transcendental claim about consciousness as the condition of possibility for appearance, including what appears to scientific observation; it is not an empirical claim that 'consciousness is the essential feature for unifying the information into object representations as part of a meaningful world'.

Nevertheless, Newen's considerations do not undermine this empirical claim, for several reasons. First, contrary to Newen, it is far from obvious that 'representations with [exactly] the same content can be processed either consciously or unconsciously'. Second, in the context of the studies involving pathology (visual agnosia, hemineglect, blindsight), 'conscious' could mean accessible to verbal report, or phenomenal but not accessible to report and action guidance. Newen's conclusion that these studies show that consciousness is not required for certain types of processing is too quick unless it carefully distinguishes between different senses of 'consciousness' and different ways that consciousness can be assessed and measured. Finally, in every case of 'unconscious' processing he mentions, the

subjects are conscious in the sense of being awake and having a unified (though disrupted) phenomenally conscious field of awareness (see pp. 351–2). So these 'unconscious' processes do not constitute a case of perceptual processing in the complete absence of consciousness.

7. Enaction and Consciousness

In this section Newen is concerned with certain aspects of Alva Noë's view of perception. Since I do not make use of those aspects in *Mind in Life*, I will not comment on them here, except to say that, to my mind, Noë (2004; 2005) has already dealt satisfactorily with the objection Newen raises. Instead, I urge Newen to read **Siewert's** commentary, which does an excellent job of showing how perceptual experience is constituted through bodily movement.

8. Conclusion

Newen thinks functionalism and identity-theory can account for all the phenomena discussed in my book, a preposterous claim considering the failure of these theories to provide any kind of comprehensive and satisfactory account of the mind, let alone the specific phenomena I discuss.

Susan Oyama

Oyama raises a number of important concerns about the relationship between the theory of autopoiesis and Developmental Systems Theory. I like her conception of these theories as 'neighbours' and I agree that my claim of 'complementarity' for the two theories still requires further working out in relation to the questions she raises concerning internalism and causal asymmetry (the determination of what counts as 'inside' versus 'outside', and the attributing of asymmetrical causal roles to internal versus external factors). I cannot fully address those concerns here, so I offer the following general remarks as a way to keep the conversation going.

In *Mind in Life*, I write that autopoiesis (in a broad sense that includes adaptivity) is the 'self-production of an inside that also specifies an outside to which it is normatively related', and thus that autopoiesis is best seen as the 'dynamic co-emergence of interiority and exteriority' (p. 79). Yet I also immediately go on to say that 'there seems to be an asymmetry here, for it is the internal self-production process that controls or regulates the system's interaction with the outside environment' (*ibid.*) To support this point, I quote two philosophers and theoretical biologists, Alvaro Moreno and Xabier Barandiaran (2004), who write

about what they call, following Varela, the 'basic autonomy' of life: 'the (self) generation of an inside is ontologically prior to the dichotomy in-out. It is the inside that generates the asymmetry and it is in relation to this inside that an outside can be established. Although the interactive processes [and] relations are necessary for the maintenance of the system, they presuppose it (the system) since it is the internal organization of the system that controls the interactive relations' (Moreno and Barandiaran, 2004, p. 17).

A number of authors, including Oyama in her commentary, have expressed worries about this assertion of asymmetry between interior and exterior. Donn Welton (2011) suspects it of being a kind of 'bio-idealism', and argues that it unduly downplays the way the environment leads the organism into certain rhythms, behaviours, and internal transformations — an environmental role he calls 'affective entrainment'. John Protevi (2010) wonders whether Varela's notion of an autonomous system 'overemphasizes the individual as self-conserving product as opposed to individuation as always ongoing process'.

I am sympathetic to these helpful and friendly (or neighbourly) criticisms, for a certain tendency to privilege interiority in autopoietic discourse has always troubled me. I felt that worry in writing those words in *Mind in Life* about the reciprocal yet asymmetrical relation between interiority and exteriority, but I did not adequately address the worry because of another argument I was trying to advance, specifically that the genuine interiority of life is a precursor to the interiority of consciousness, and hence that the conception of nature presupposed in standard formulations of the hard problem or explanatory gap for consciousness — namely, that living nature has no genuine interiority — is misguided.

Here is another way to come at the issue about interiority Oyama raises. On the one hand, I claim that the adaptive-autopoietic process 'brings forth' or 'enacts' what counts as the living being's world, and not the reverse; on the other hand, I claim that the living being and its environment are 'structurally coupled', and interiority and exteriority are 'dynamically co-emergent'. So how do we resolve this issue of asymmetry in the reciprocal coupling of living beings and their worlds?

At this point I would like to inject an autobiographical remark to indicate how long this tension has preoccupied me. Varela and I began working together on *The Embodied Mind* in the late 1980s when I was a graduate student. It was during those years that Varela introduced into his work the terminology of organisms 'enacting' and 'bringing forth' their worlds, rather than representing them (though this idea

was already implicit in his earlier work on autopoiesis with Maturana). This way of talking worried me — precisely for its not fully worked-out suggestion of some kind of idealism or constructivism. So whenever Varela would write that the organism enacts its world, I would try to rewrite the sentence to say that a world is brought forth or enacted by the structural coupling of the organism and its environment. My aim was to shift the emphasis away from the organism as the enactor of its world to the relational process of enactment. Varela was happy with these changes, as they fitted better our other sympathies with the Indian Buddhist concept of dependent co-origination (*pratītyasamutpāda*), also central to *The Embodied Mind*. Nevertheless, my re-wording clearly did not deal adequately with the tension, for the question of the asymmetrical versus symmetrical status of the organism — or of the adaptive autopoietic process — in the relational process of enactment remained unanswered.

Welton (2011) proposes a way to resolve this tension with his notion of affective entrainment. He writes, referring specifically to the requirement of adaptivity for sense-making:

> Adaptation is much more than a dynamic adjustment allowing the organism to get along better with its habitat according to internal self-generated norms. It is also a transformation of the organism's internal processes and norms according to the demands of an environment that introduces 'sense-producing' or 'sense-demanding' requirements of its own… [T]he environment that the organism opens or *enacts* is also the world that *entrains* it and reflexively transforms both the processes and the structure of the cell 'reacting' to it. (Welton, 2011)

In the case of bacteria, for example, the presence of sucrose exerts an 'extrinsic' control over chemotaxis; in dynamical systems language, sucrose acts as an external control parameter, entraining the cells to swim up-gradient. It is precisely this entrainment, Welton maintains, that accounts for the status of sucrose as attractant.

I welcome and agree wholeheartedly with these points. Living as sense-making is systemically generated and, as Oyama notes, systemically controlled. Living beings enact environments that pull them along into certain rhythms, behaviours, and internal transformations (this point becomes especially important when we remember that the environment is always an environment of other living beings — bacteria, for example, do not live in isolation but in microbial communities). In Welton's words: '*The organism enacts an environment as the environment entrains the organism*. Both are *necessary* and neither, by itself, is *sufficient* for the process of sense-making' (*ibid.*)

But now comes the tricky point. What we have just said implies that the relation between organism and environment is *reciprocal*, for each acts as a control parameter for the other. But this kind of reciprocity does not imply that their relation is not also *asymmetrical*, in the relevant sense of asymmetry. Although the physical and energetic coupling between a living being and the physico-chemical environment is symmetrical, with each partner exerting more influence on the other at different times, the living being typically modulates the parameters of this coupling in a way the environment typically does not (Barandiaran *et al.*, 2009). Living beings, precisely because they are autonomous and adaptive, can 'surf' environmental events and modulate them to their own ends, like a bird gliding on the wind. 'Interactional asymmetry' is precisely this capacity to modulate the coupling with the environment. If we lose sight of this interactional asymmetry, then we lose the ability to account for the directedness proper to living beings in their sense-making, and hence we lose the resources we need to connect sense-making to intentionality.

It is crucially important, however, to realize that 'boundary' in this context cannot be identified with any given spatial boundary, such as a membrane, but refers instead to the system's topological boundaries as an autonomous network of processes. The way these processes are structurally realized is plastic, both compositionally (what materially composes them over time) and spatially (where they are located in relation to spatially specified boundaries). The processes constitutive of an autonomous network can incorporate external material resources and extend beyond the biological membrane of the body, as happens, for example, when a blind person uses a cane to perceive the environment (see Di Paolo, 2009; Thompson and Stapleton, 2009). Thus Gregory Mengel (as reported by Oyama) gets me right when he says that my 'internal-external distinctions are less about spatial boundaries… than about selfhood, organizational closure, and the context-dependence of causes'. Furthermore, as far as I can see, interactional asymmetry in the above sense does not involve the kind of arbitrary causal privileging that Developmental Systems Theory criticizes, because such asymmetry does not accord a special causal status to processes just because they happen to occur on one or the other side of some spatial boundary.

John Protevi

I find the links Protevi makes between my project in *Mind in Life* and Deleuze's writings fascinating, but I do not know Deleuze well enough

to respond, so I will focus on what he calls the 'question of panpsychism' (see also **Wheeler**, who raises the question of panpsychism).

Protevi thinks that although my conception of the deep continuity of life and mind escapes from the Cartesian problem of the relation between the mental and the physical, it raises the problem of the emergence of life and mind from non-life. He wonders whether I am too restrictive in my conception of mind when I trace mind back to living as sense-making. Moreover, given that I work with the concepts of processes and networks as webs of processes, what is to stop me from embracing the kind of process panpsychism we find in Whitehead or Deleuze?

To address this issue I want to compare Jonas and Merleau-Ponty, because it is precisely on this issue about matter and life — or what Merleau-Ponty (in *The Structure of Behavior*) calls the physical order and the vital order — that I follow Merleau-Ponty and not Jonas.

Jonas contrasts the *wave* and the *organism*. The wave he takes to be a material aggregate, which, as 'an integrated event-structure', has no ontologically emergent status. He writes that to the wave 'no special reality is accorded that is not contained in, and deducible from, the conjoint reality of the participating, more elementary events'. In other words, Jonas accepts analytical and ontological reductionism for physical phenomena. What he then argues is that this kind of reductionism fails in the case of the organism, which *is* ontologically emergent. Life, as he puts it, is thus an 'ontological surprise'.

Now, if we follow this line of thought, then I think we do face a serious life–matter problem, analogous to the mind–body problem. How does life emerge from non-life? The panpsychist argues that we cannot make good on this invocation of emergence, that it is ultimately mysterious. Hence the options would seem to be either some kind of dualism or some kind of panpsychism.

But this line of thought is not at all the one we find in Merleau-Ponty. Already in *The Structure of Behavior*, Merleau-Ponty rejects analytical reductionism for physical forms like waves, soap bubbles, and convection rolls (see pp. 72–3). As he writes, 'The genesis of the whole by composition of the parts is fictitious. It arbitrarily breaks the chain of reciprocal determinations' (Merleau-Ponty, 1963, p. 50). Consider also this passage, which I quote in *Mind in Life* (p. 72):

> [E]ach local change in a [physical] form will be translated by a redistribution of forces which assures us of the constancy of their relation; it is this internal circulation which is the system as a physical reality. And it is no more composed of parts which can be distinguished in it than a melody (always transposable) is made of the particular notes which are

> its momentary expression. Possessing internal unity inscribed in a segment of space and resisting deformation from external influences by its circular causality, the physical form is an individual. It can happen that, submitted to external forces which increase and decrease in a continuous manner, the system, beyond a certain threshold, redistributes its own forces in a qualitatively different order which is nevertheless only another expression of its immanent law. Thus, with form, a principle of discontinuity is introduced and the conditions for a development by leaps or crises, for an event or for a history, are given. (Merleau-Ponty, 1963, p. 137)

As I state in *Mind in Life*, this description of physical form as introducing a principle of discontinuity and the conditions for development by 'crises' has been borne out by René Thom's 'catastrophe theory', which mathematically describes abrupt transitions and qualitative discontinuities in physical systems, and by Jean Petitot's extension of Thom's work to a morphodynamical 'physics of phenomenality', which aims to bridge the gap between the microphysical substrate and macrophysical forms.

What we find in Merleau-Ponty is a reconceptualization of matter, life, and mind, one that does not bring mind down into the domain of microphysical processes nor equate mind with information transfer and self-organization, but rather tries to show how the notion of form as dynamic pattern or individuation process can both integrate or bridge the orders of matter, life, and mind, while also accounting for the originality of each order. This is the path I try to follow in *Mind in Life* and not panpsychism.

Nevertheless, I admit that my characterization in *Mind in Life* of life as autopoiesis plus cognition can be read as equating mind and life, and hence as opening a door to the panpsychist line of thought. What I would now rather say — and these remarks pertain also to issues raised by **Newen** and by **Wheeler** — is that living is sense-making and that cognition is a kind of sense-making. A wave or a soap bubble is an individuating process but not a sense-making one, because it does not modulate its coupling with the environment in relation to virtual conditions and norms. A unicellular organism is a self-individuating and sense-making being but not a cognitive one if by 'cognition' we mean being intentionally directed toward objects as unities-in-manifolds having internal and external horizons (call this the phenomenological sense of cognition). Ultimately, what matters to me is not to fix the meanings of the words or concepts 'matter', 'life', 'mind', 'cognition', and so on. Instead, my aim is to see whether we can chart multiple

passages back and forth between those orders that we conceptualize — in different ways and at different times — as matter, life, and mind.

Charles Siewert

Siewert focuses on my treatment of the explanatory gap. He raises a challenge for my treatment and develops an account of embodied consciousness in order to meet the challenge. I welcome this account and think it takes an important forward step that builds on what I write in Chapter Eight of *Mind in Life*.

Siewert finds a lacuna in my argument that a proper phenomenological account of perceptual experience shows that the same perceptual functions that occur in our world could not occur in a world in which there were no experience. I argue that it is on the basis of our kinaesthetic experience of our own body that we are able to perceive objects in space as unities in and through perspectivally varying appearances, and thus that bodily self-experience is constitutive of the perceptual function of individuating objects in space. This perceptual function, I argue, would not occur in a world in which there were no bodily self-experience; hence there could not be a zombie that was functionally equivalent to us in its perceptual abilities. Siewert wonders, however, whether I have not stipulatively built a tie to bodily self-experience into the concept of perceptual functioning. My claim, however, is that our perceptual abilities to individuate and track objects in space depend constitutively on bodily self-experience; it is thus a claim about the functional role consciousness plays in perception. In making that claim, I do not, as far as I can see, build bodily self-experience into the concept of the relevant perceptual functions. In any case, the lacuna Siewert sees is the need for a proper demonstration that perceptual functioning 'cannot be factored into a "phenomenal, experiential bit" and a "bodily movement bit" with no more than a contingent causal link between them'. I agree that I have not provided such a demonstration, and I am grateful to Siewert for working to provide it.

Siewert presents an analysis of perceptual experience that shows that looking and touching are indissolubly both consciousness and movement, and hence that our bodily engagement with the world cannot be factored into a phenomenal component and a movement component that are only causally and contingently related. I greatly admire Siewert's analysis. Had I had thought of it myself I certainly would have included it in order to bolster my argument that perceptual experience is so tied to embodied activity that the same perceptual

functions occurring in our world could not occur in a world in which there were no experience. Siewert's defence of the 'Embodied Consciousness View' (Section 3) and his application of this view to the explanatory gap are very helpful additions to my efforts.

I also admire the way Siewert locates the embodied-consciousness view in relation to other views of the explanatory gap. The only place where we seem to differ concerns zombies. Siewert distinguishes between a zombie world functionally equivalent to ours and a zombie world physically equivalent to ours. My 'phenomenological critique' of zombies targets the functional version but does not show that a physical zombie world is inconceivable. Nevertheless, it is far from clear that a physical zombie world *is* conceivable. To believe that such a world is conceivable we need to know how to imagine this world in such a way that arbitrary details about it can be filled in without any incoherence arising. To my knowledge, no proponent of zombie conceivability ever provides this kind of guidance. One of my aims in Chapter Eight of *Mind in Life* is to show how unsatisfactory this way of thinking is, especially when it is tied — as it typically is — to phenomenologically impoverished ways of thinking about consciousness in relation to embodiment (see pp. 233–4).

Siewert also raises a question at the end of his commentary. The question concerns how the 'inwardness' of biological life relates to the inwardness of subjectivity. He wonders whether inwardness in the sense of maintaining a boundary between inside and outside is necessary for inwardness in the sense of being a suitable focus of empathic regard, and he asks whether I mean to show how the first can be built up to the second, and whether we need to close that gap in order to explain consciousness.

My remarks here will be brief; for related considerations see my replies to **Newen**, **Protevi**, and **Wheeler**. My view is that life or living being already involves a kind of inwardness that goes beyond what can be captured in an external conception of material structure and mechanical function. That kind of inwardness is not mere spatial inwardness (as marked by a membrane) but rather the inwardness of immanent purposiveness and being normatively related to the environment. I maintain that this kind of inwardness is a precursor to subjectivity, but I do not show how the one can be built up to the other; that task requires much more work and belongs to what I call the 'body–body problem'. A full account of consciousness and its relation to biological life would require an explanatory bridge from the one kind of inwardness to the other, but there is plenty of work that can be done on consciousness without that kind of bridge being in place, as

Siewert's own presentation of the embodied-consciousness view amply demonstrates.

Robert Van Gulick

Like **Dennett**, Van Gulick sees my rhetoric in *Mind in Life* as 'radical' or 'iconoclast' in relation to the 'mainstream literature', which I 'attack'. For the reasons indicated in my response to Dennett, I do not share this way of looking at my efforts in relation to the complex terrain of philosophy of mind, cognitive science, and theoretical biology. Although my views are certainly iconoclastic from the perspectives of functionalism, cognitivism, and adaptationism, I prefer to think of myself as a pluralist who draws from the many approaches that populate contemporary science and philosophy.

Van Gulick offers teleofunctionalism and biosemantic theories of meaning as approaches that share many of my ideas about the deep continuity of mind and life. At a general level, I agree, but this resemblance strikes me as superficial. There are two crucial differences between these theories and my approach (see also **Hutto**). First, these theories treat organisms as heteronomous systems, not as autonomous ones; and second, these theories assume an adaptationist account of evolution instead of the enactive one I propose.

Van Gulick thinks it odd that I do not mention non-reductive physicalism in my list of mainstream options for dealing with the explanatory gap. But non-reductive physicalism is arguably an unstable position that ultimately results in either eliminativism or property dualism. So it does not provide a genuine alternative to either dualism or materialism. As for my position in relation to other approaches to the explanatory gap, see **Siewert** for a very useful assessment of the lay of the land.

Van Gulick cites Lycan's teleofunctionalist approach as convergent with mine. In the Lycan passage Van Gulick quotes, however, Lycan proposes a reductionistic account that would explicate the mental in non-mental terms through a strategy of homuncular decomposition. My approach is very different. First, I do not believe that mental terms can be reductively translated into non-mental terms. Second, I maintain that homuncular decomposition fails for complex systems (see Appendix B). Third, Lycan's conception of function is extrinsic function, not the immanent purposiveness of autonomous organization (see pp. 144–6).

There is also a much larger background philosophical difference between these theories and my approach. Non-reductive physicalism,

biosemantics, and teleofunctionalism advance a kind of naturalism that denies the transcendental status of consciousness and subjectivity. The kind of naturalism I propose, however, recognizes the transcendental status of consciousness and reconceptualizes nature from a transcendental standpoint.

Van Gulick states that few if any ontological physicalists subscribe to mereological reductionism because they assert that 'the properties of wholes are fully determined by the intrinsic properties of their basic parts plus their mode of composition and arrangement'. Like nonreductive physicalism, however, this position is arguably unstable, and leads either to mereological reductionism or to property dualism. My account of emergence is meant to offer a third way between these problematic alternatives (see Appendix B).

Michael Wheeler[1]

Wheeler raises a number of important questions about the 'deep continuity thesis of life and mind' through a probing analysis of the relations among the concepts of autopoiesis, autonomy, adaptivity, sense-making, cognition, and teleology. I am grateful for this analysis because it gives me the opportunity to formulate the interconnections among these concepts more clearly than I did in *Mind in Life*.

Let me start with the following schematic presentation. I maintain the following theses:

(1) Autopoiesis and adaptivity are individually necessary and jointly sufficient for life. In other words, life is autopoiesis plus adaptivity.
(2) Autonomy and adaptivity are individually necessary and jointly sufficient for immanent purposiveness (each part being both a product and producer of the other parts, so that the whole system is a self-organizing whole) and sense-making (behaviour or conduct in relation to significance, valence, and norms that the system itself brings forth or enacts on the basis of its autonomy).
(3) Sense-making is cognition in a wide sense of the term; or, to put the point another way, sense-making is the basic mark of the cognitive.
(4) Autopoiesis is the paradigm case of autonomy, in the sense that it is the best understood case and the minimal case of an autonomous organization (more on 'minimal' below).
(5) Thus, autopoiesis and adaptivity are jointly sufficient for immanent purposiveness and sense-making.

---

[1] I wish to thank Ezequiel Di Paolo for helpful discussion of Wheeler's commentary.

(6) Thus, life is autopoiesis plus cognition (any living system is an adaptive autopoietic system, hence an adaptive autonomous system, hence a sense-making system, hence a cognitive system).
(7) Thus, cognition is necessary for life.

We thus arrive at one sense of the deep continuity thesis of life and mind: *wherever there is life there is mind*.

The crucial remaining question is whether autopoiesis is necessary for autonomy and hence also necessary for sense-making. If the answer to this question were yes, then autopoiesis would be necessary for cognition, and we would arrive at another sense of the deep continuity thesis, namely, that life is necessary for mind, or *wherever there is mind there is life*.

To answer this question we need to be clear about what autopoiesis requires. Wheeler states that 'autopoiesis is autonomy plus materiality', by which he means that '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'. Although this description is true, it does not specify the crucial definitive feature of the *autopoietic organization* that makes an autopoietic system a particular kind of autonomous system. What makes the system autopoietic is not its self-produced material boundedness as such, but rather that the relations constituting the system are relations between processes of molecular transformation, including those that make up the boundary. To put the point another way, the *physical* boundary of a cell should not be confused with its *organizational* boundary (see also my replies to **Dennett**, **Newen**, and **Oyama**). A molecule traversing the cell membrane is spatially inside the boundary that contains most of the processes that make up the autopoietic system, but this location does not by itself determine whether the molecule is part of the autopoietic system. The molecule can become part of the autopoietic system only if it can become integrated into the network of processes that sustain the relation of closure among themselves (e.g. if the molecule becomes a participant in metabolic reactions). If the molecule does not participate in such processes or if it disrupts them, then the molecule remains 'foreign' (non-self).

When we ask whether autopoiesis is necessary for autonomy and sense-making, it is crucial to realize that what we are asking is whether the autopoietic organization is necessary for autonomy and sense-making, not whether the 'dual materiality of the systemic

boundary and the systemic mode of existence' is necessary for autonomy and sense-making.

In addition, we need to refine what 'necessary' means. I maintain that there are autonomous systems that are not themselves first-order autopoietic systems, such as immune networks, the nervous system, insect colonies, animal societies, and primate bands. Multicellular organisms are a tricky case (see pp. 105–7); so too is Gaia (the Earth's ecosphere). In any case, in one important sense, I maintain that autopoiesis is not necessary for autonomy (*viz.* it is not necessary that an autonomous system itself be autopoietic). Notice, however, that in all these cases, the autonomous system depends constitutively on having autopoietic components.

Wheeler is thus doubly misguided, I think, when he writes 'such materiality [belonging to autopoiesis] is apparently expendable by the time that biology gets as far as the nervous system'. On the one hand, as we have seen, it is not the materiality *per se* of an autopoietic system that is the relevant feature, but rather the way the system realizes the autopoietic organization in the molecular domain. In the case of the nervous system, the system depends constitutively on components (nerve cells and glial cells) that realize the autopoietic organization in the molecular domain. On the other hand, it is far from obvious and arguably false that the materiality of the nervous system is expendable (think of how the electrical signalling properties of neurons are chemically realized, and how the nervous system depends crucially for its functioning on complex molecular cross-talk with the immune and endocrine systems).

Wheeler might respond that the materiality of the nervous system is expendable in the sense that the cognitive functions the nervous system implements could be realized in a materially different structure. This thesis of 'multiple realizability', however, should be treated as an empirical hypothesis that could well be false (Shapiro, 2004). Here we need to distinguish between the following two ideas — *multiple realizability* and *compositional plasticity*. If a functional property (e.g. being a watch) can be implemented in different physical mechanisms with different causal properties (analogue watch, digital watch, sundial), then that property is multiply realizable. Mere difference in physical composition, however, is not sufficient for multiple realizability; the compositional difference must entail a difference at the level of mechanisms and causal properties. Suppose we constructed an artificial nervous system. If the artificial 'neurons' realize functional states by virtue of the same electrical properties as biological neurons, then the artificial brain does not realize the functional states

in a relevantly different way (Shapiro, 2004). We can thus distinguish between the following two empirical hypotheses: (i) the *multiple realizability* hypothesis: the cognitive functions the nervous system implements can be realized in different physical systems with different causal properties; (ii) the *embodied hypothesis*: the cognitive functions the nervous system implements can be realized only in systems having the causal properties of the biological nervous system. According to the first hypothesis, the materiality of the nervous system is expendable; according to the second it is not. Deciding between these two hypotheses requires evaluating the empirical evidence — it cannot be decided on the basis of conceptual considerations alone.

We thus arrive at another sense in which we can ask whether autopoiesis is necessary for autonomy, namely, whether autopoiesis is a necessary ingredient for autonomy, or to put it another way, whether there could be an autonomous system that did not realize its autonomy through autopoietic constituents and thus did not depend constitutively on autopoiesis.

Before addressing this question, I need to clear up a few points arising from Wheeler's reading of my views.

First, when I speak of 'minimal autonomy' or 'basic autonomy', I am referring to the simplest systems we know that have all the required properties for autonomy. The paradigm case is the autopoietic cell. Other autonomous systems (e.g. multicellular organisms or insect colonies) are non-minimal because they are more complex and not found historically before the appearance of the minimal versions.

Second, the passage Wheeler quotes from page 160, where I write, 'minimal autonomy *depends* on macromolecules but requires that these macromolecules be *organized* in a particular way, namely, the autopoietic way', occurs in the context of a disagreement with Dennett (see my response to **Dennett** above). My point is that, contrary to Dennett, replicating macromolecules such as DNA/RNA do not meet the conditions required for minimal autonomy (in the sense just indicated), and hence do not meet the conditions required for agency; they lack the proper organization, namely, the autopoietic organization.

Finally, it is not the case that 'the only thing that autopoiesis adds to the concept of autonomy is the dual materiality of the systemic boundary and systemic mode of existence'. Following Varela (1979), I define an autonomous system as one in which the constituent processes (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 (see p. 44). This specification leaves entirely open the processes that can be interrelated in this way. What autopoiesis adds to this specification is that the processes are ones that modulate molecular transformations in the chemical domain. To repeat a point I made earlier, the crucial feature of such a biochemical instantiation of autonomy is not the dual materiality *per se* but the fact that this materiality realizes a certain organizational boundary. Put another way, the identity of an autopoietic system cannot be defined by preservation of the membrane (despite such preservation being critical for the maintenance of autopoiesis); it has to be defined by preservation of the network's organizational boundary, for it is this boundary that constitutes the crucial interface demarcating the system from its world.

I come now to the question of whether autonomy and sense-making require autopoiesis, in the sense of depending constitutively on autopoiesis. Here I admit to being unsure, for the reasons already sketched in my reply to **Newen**. Let me now spell out those reasons in more detail.

On the one hand, as work on autonomous systems in AI and robotics suggests (see Barandiaran *et al.*, 2009; Froese and Ziemke, 2009), it seems conceivable that there could be an adaptive self-constituting system that was not based on autopoietic constituents. For example, perhaps it is possible to bypass autopoiesis and construct directly a sensorimotor agent that achieves its autonomy at the level of an adaptive and organizationally closed sensorimotor loop.

Notice, however, that for the system to be genuinely autonomous, it would need to (i) be an individual, in the sense of continually enacting or bringing forth its own existence in challenging thermodynamic conditions (where 'its own existence' or its individuality is defined by its topology and organization as a network, not its material or spatial boundedness *per se*); (ii) be the active source of its interactions, in the sense of modulating the parameters of its coupling with the environment on the basis of its internal (self-organized) activity ('interactional asymmetry'); and (iii) generate the norms for those interactions on the basis of its activity ('normativity') (Barandiaran *et al.*, 2009). No existing robot (e.g. Rodney Brooks 'Creatures') meets these criteria.

On the other hand, given these criteria it is not unreasonable to doubt that they can be achieved without autopoietic constituents. These criteria require, at the sensorimotor level, not just a self-constituting system embedded in sensorimotor interactions, but a self-constituting system that can adaptively regulate its sensorimotor interactions. As Froese

and Ziemke (2009) argue in a careful and compelling analysis, such a system 'must bring forth its sensors, effectors, and their internal organizational link (some adaptive mechanism) on the basis of its self-constituting operations. So far, no one has been able to artificially generate such a system' (*ibid.*, p. 495). Moreover — and this is the crucial point for my purposes here — it is hard to see how this requirement could be met without something like a *metabolism*. Put another way, it is hard to see how this requirement could be met *without something like an autopoietic organization for the constituents that make up the sensors, effectors, and the adaptive mechanism that links them*. Notice that such an autopoietic organization need not be materially realized in the same organic way as our terrestrial cells; in that sense, both the autonomous and autopoietic organizations are *compositionally plastic*. Nevertheless, on the present line of thought, autonomy would not admit of *multiple realizability*, in the sense of being implementable in non-autopoietic mechanisms having causal properties different from those of autopoiesis.

We thus come back to the second sense of the deep continuity thesis, namely that autonomy and sense-making require adaptive autopoiesis, understood to mean that autonomy and sense-making depend constitutively on adaptive autopoiesis. If this thesis is true, then life is necessary is for mind.

This second sense of the deep continuity thesis converges with a third sense of the thesis, which I will call the existential-phenomenological sense of the deep continuity of life and mind. According to this thesis, which derives from the writings of Hans Jonas, certain existential structures of human life or phenomenological structures of human experience — notably, self/world, freedom/necessity, being/not-being — are applicable to life itself (pp. 129, 157).

Here is how the convergence between the second and third versions of deep continuity goes. Jonas traces selfhood and the having of values, purposes, and norms down to the basic phenomenon of metabolism: 'to an entity that carries on its existence by way of constant regenerative activity we impute *concern*. The minimum concern is to be, i.e., to carry on being' (Jonas, 1968). Jonas argues that without metabolic self-construction, there would be no such thing as the constitution of a meaningful perspective by a system for that system. This idea provides the existentialist side of the deep continuity thesis: without constant self-construction or self-creation in a finite, contingent, and challenging environment — in what Ezequiel Di Paolo (2009) calls 'precarious conditions' — there would be no such thing as subjectivity marked by the polarities of self/world, immanence/

transcendence, freedom/necessity, and being/not-being. In this way, Jonas's existential-phenomenological analysis of life converges with the line of thought that suggests that adaptive autonomy and sense-making depend constitutively on autopoiesis. In short, both lead to the conclusion that life is necessary for mind.

If mind is necessary for life and life is necessary for mind, then Wheeler is right that 'one might as well say that life is identical with cognition' (or that life is identical with mind). Like **Protevi**, Wheeler thinks this equation results in something like panpsychism, though, unlike Protevi, who favours panpsychism, Wheeler thinks it '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'.

Let me repeat here my response to Protevi. On the one hand, I am indeed willing to bite the bullet and say that life is mind and mind is life (though with the reservation that I remain unsure about whether autonomy requires autopoiesis). On the other hand, if I were writing *Mind in Life* today, I would choose a different formulation. I would say that living is sense-making and that cognition is a kind of sense-making. But I would do so for the sake of terminological and conceptual clarity, not in order to mark some clean break in nature. My aim would be to mark the difference between sense-making as such (comportment in relation to significance and norms), and the kind of sense-making that requires intentionality in the proper phenomenological sense — intuitive intentionality (empty and filled intentions in perception, memory, and imagination), signitive intentionality (pictures, signs, indications), and categorial intentionality (propositional and conceptual thought).

I am now in position to say something about Wheeler's discussion of what 'enrichment' means when I write (following Jonas), 'certain existential structures of human life are an enriched version of those constitutive of all life' (p. 157), or more simply put, that human being is an enriched version of living being.

In my view, enrichment does not mean that something gets added from outside life in order to make it mind. Instead, life evolves in such a way as to transform sense-making, and as this evolution happens the more complex forms of sense-making reach back, as it were, and transform the simpler ones. There are many ways to tell this story, but a Jonas way to tell it would be to say that the self-identity and world-relatedness of living being evolves into increasingly complex forms of self-identity and increasingly mediated forms of relating to the world. On the one hand, the simpler forms are preserved in the more complex; on the other hand, the more complex forms alter the simpler

ones and make them dependent on the complex ones. For example, human culture penetrates virtually every aspect of our metabolism, so that there is no such thing as 'naked' human metabolic being independent of our cultural ways of living. Enrichment is thus never a mere addition but always an overall transformation of life and mind. For this reason, even if it should turn out that life is not necessary for mind (that autonomy and sense-making do not require autopoiesis), it would still be the case, contrary to what Wheeler says at the end of his commentary, that mind would be *in* life and not simply life in mind.

Wheeler raises one other large issue in his commentary — the relation of the deep continuity thesis of life and mind to the extended cognition hypothesis. I have discussed this issue elsewhere (Thompson and Stapleton, 2009), and do not have the space to present a full discussion here, so I will make only a few points.

Nothing in my view prevents me from allowing that there can be immanently purposive systems that incorporate elements whose function is specified extrinsically (see Thompson and Stapleton, 2009). Think of a prosthetic limb that is incorporated into a person's ongoing life. For a system to be immanently purposive it is not necessary that every element that participates in the system be materially produced by that system. Furthermore, immanent purposiveness does not mean that the parts must produce each other in the autopoietic sense; it means that they must generate and realize themselves as a whole according to the definition of autonomy. Finally, it remains an open question whether immanent purposiveness depends constitutively on autopoiesis, or whether there can be immanent purposiveness without autopoietic constituents. For these reasons, I see no *inconsistency* between my deep continuity view and the extended cognition hypothesis.

Nevertheless, there are significant *tensions* between the two approaches. One tension arises from the way the extended cognition hypothesis discusses cognition as spatially located versus the way the enactive approach treats cognition as relational. Wheeler comments on this issue in note 7, but seems to miss the point. Of course, relations can exhibit spatiality in the sense that the terms of a relation (e.g. laptop, table) are themselves spatially located. But where is the relation (on top of) itself located? It is something like a category mistake to think that spatial relations are themselves spatially located in the way the terms of the relation are. Similarly, the point both Thompson and Stapleton (2009) and Di Paolo (2009) are making is that it does not make sense to think of cognition as spatially located in the way that the 'vehicles' enabling cognitive processes are spatially located.

Dan Zahavi

Zahavi focuses on my book in relation to the recent debate about the possibility of 'naturalizing phenomenology'. He raises two questions. The first asks whether and how 'analyses pertaining to subpersonal processes and mechanisms can possibly influence and enrich phenomenological accounts that attempt to do justice to the first-person perspective and seek to understand the experience in terms of the meaning it has for the subject'. The second asks 'how deeply committed' I am to transcendental thought. For example, do I 'endorse some kind of compatibility between empirical realism and transcendental idealism'?

The first question Zahavi winds up answering himself. We might start with a certain phenomenological description and then revise or enrich this description on the basis of empirical investigation. For example, a number of traditions in western philosophy and psychology distinguish between reason and passion, or cognition and emotion. Some processes are thought to be purely cognitive and others purely affective. Recent experimental psychology and cognitive neuroscience, however, strongly speak against this view. Many behavioural and neuroscientific findings indicate that there is no separation between cognition and emotion: every cognitive process is also affective, and every brain area traditionally described as cognitive also belongs to emotion and *vice-versa* (see Colombetti and Thompson, 2007; Pessoa, 2008; and Chapter Twelve of *Mind in Life*). These discoveries at neural and behavioural levels can and should provoke a phenomenological re-examination of experience at the personal level. What might at first have seemed separate and only instrumentally related processes in experience (for example, an emotion and a reflective judgment) can be shown to be constitutively interdependent instead.

Nevertheless, I agree with Zahavi that 'the discovery of a significant complexity at a subpersonal level… cannot by itself force us to refine or revise our phenomenological description. It can only serve as a motivation for further enquiry'. There is one qualification, however, that I would add to this remark. Although fMRI or EEG evidence on its own counts as evidence about only subpersonal neural correlates, when such evidence is linked to behaviour and self-report it is no longer purely subpersonal. Instead, it encompasses the personal level as probed from third-person and second-person perspectives. Motivations from psychology and neuroscience to refine and revise our phenomenological descriptions do not come from the strictly subpersonal

level; they come from the way this level is systematically related to the personal level in experimental investigation.

I also agree with Zahavi that these points are underdeveloped in *Mind in Life*. It remains for future work to develop them more systematically and with a greater number of examples.

The question of how I see transcendental philosophy in relation to empirical science is a large issue that I cannot deal with satisfactorily here, so I will limit myself to a few points.

The foundation of transcendental thinking is rigorous and careful attention to *how* things are given to experience. Every claim about *what* something is presupposes that the object of this claim is given in some *way* to our experience, where 'experience' is taken widely to include thought, and where what counts as a 'way of experiencing' cannot be absolutely delimited in advance. Such attention to how things show up, to the ways they present themselves, necessitates a critical and reflexive stance toward any cognitive claim, a stance that requires us to take account of the standpoint of the cognizer in making that claim. In this way, we are lead back to consider subjectivity and intersubjectivity as conditions of knowledge and knowledge production.

This commitment has always been a central component of the enactive approach (see *The Embodied Mind*, Chapter One); it is also what is meant by Maturana and Varela's statement, 'everything said is said by an observer' (Maturana and Varela, 1987).

Now, this basic commitment to critical reflexivity can be developed in a variety of different ways and to many different ends. In the case of transcendental phenomenology, it leads to the analysis of phenomena such as intentionality, the lived body, intersubjectivity, and time consciousness. In my view, all these analyses — including the ones that lead Husserl to embrace what he means by transcendental idealism — are compatible with empirical realism. As Zahavi notes, however, and as I also maintain, the systematic development of these analyses do not leave empirical realism unchanged, for they force us to rethink the concept of nature in ways that can also lead to a transformation of natural science.

References

Bachman, P.A., *et al.* (1992) Autocatalytic self-replicating micelles as models for prebiotic structures, *Nature*, **357**, pp. 57–59.

Barandiaran, X., *et al.* (2009) Defining agency: Individuality, normativity, asymmetry and spatio-temporality in action, *Adaptive Behavior*, **17**, pp. 367–386.

Bourgine, P. & Stewart, J. (2004) Autopoiesis and cognition, *Artificial Life*, **20**, pp. 327–345.

Christoff, K., *et al.* (2009) Experience sampling during fMRI reveals default network and executive system contributions to mind wandering, *Proceedings of the National Academy of Sciences USA*, **106**, pp. 8719–8724.
Clark, A. (2001) *Mindware: An Introduction to the Philosophy of Cognitive Science*, New York and Oxford: Oxford University Press.
Colombetti, G. & Thompson, E. (2007) The feeling body: Toward an enactive approach to emotion, in Overton, W., Mueller, U. & Newman, J. (eds.) *Body in Mind, Mind in Body: Developmental Perspectives on Embodiment and Consciousness*, Hillsdale, NJ: Lawrence Erlbaum Associates.
Cosmelli, D., *et al.* (2007) Neurodynamical approaches to consciousness, in Zelazo, P.D., Moscovitch, M. & Thompson, E. (eds.) *The Cambridge Handbook of Consciousness*, pp. 731–774, Cambridge: Cambridge University Press.
Dawkins, R. (1995) *River out of Eden: A Darwinian View of Life*, New York: Basic Books.
Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little Brown.
Dennett, D.C. (1992) Review of F.J. Varela, E. Thompson and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, and G. Edelman, *Bright Air, Brilliant Fire: On the Matter of Mind*, *New Scientist*, 13 June, pp. 48–49.
Dennett, D.C. (1993) Review of F.J. Varela, E. Thompson and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, *American Journal of Psychology*, **106**, pp. 121–126.
Di Paolo, E.A. (2009) Extended life, *Topoi*, **28**, pp. 9–21.
Farb, N.A.S., *et al.* (2007) Attending to the present: Mindfulness meditation reveals distinct neural modes of self-reference, *Social Cognitive Affective Neuroscience*, **2**, pp. 313–322.
Froese, T. & Ziemke, T. (2009) Enactive artificial intelligence: Investigating the systemic organization of life and mind, *Artificial Intelligence*, **173**, pp. 466–500.
Gallagher, S. & Meltzoff, A. (1996) The earliest sense of self and others: Merleau-Ponty and recent developmental studies, *Philosophical Psychology*, **9**, pp. 213–236.
Jonas, H. (1966) *The Phenomenon of Life: Toward a Philosophical Biology*, Chicago, IL: University of Chicago Press.
Jonas, H. (1968) Biological foundations of individuality, *International Philosophical Quarterly*, **8**, pp. 231–251.
Kelso, J.A.S. (1995) *Dynamic Patterns: The Self-Organization of Brain and Behavior*, Cambridge, MA: MIT Press/A Bradford Book.
Kelso, J.A.S. & Kay, B.A. (1987) Information and control: A macroscopic analysis of perception-action coupling, in Heuer, H. & Sanders, A.F. (eds.) *Perspectives on Perception and Action*, pp. 3–32, Hillsdale, NJ: Lawrence Erlbaum Associates.
Luisi, P.L. (2010) *The Emergence of Life: From Chemical Origins to Synthetic Biology*, Cambridge: Cambridge University Press.
Lutz, A., *et al.* (2002) Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task, *Proceedings of the National Academy of Sciences USA*, **99**, pp. 1586–1591.
Lutz, A., *et al.* (2008a) Attention regulation and monitoring in meditation, *Trends in Cognitive Sciences*, **12**, pp. 163–169.
Lutz, A., *et al.* (2008b) Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise, *PLoS ONE*, **3** (3), e1897, DOI:10.1371/journal.pone

Maturana, H.R. & Varela, F.J. (1987) *The Tree of Knowledge: The Biological Roots of Human Understanding*, Boston, MA: Shambala Press/New Science Library.
McMullin, B. (2004) Thirty years of computational autopoiesis: A review, *Artificial Life*, **10**, pp. 277–295.
Merleau-Ponty, M. (1963) *The Structure of Behavior*, Fisher, A. (trans.), Pittsburgh, PA: Dusquene University Press.
Moreno, A. & Barandiaran, X. (2004) A naturalized account of the inside-outside dichotomy, *Philosophica*, **73**, pp. 11–26.
Nielsen, L. & Kaszniak, A. (2006) Awareness of subtle emotional feelings: A comparison of long-term meditators and nonmeditators, *Emotion*, **6**, pp. 392–405.
Noë, A. (2004) *Action in Perception*, Cambridge, MA: MIT Press.
Noë, A. (2005) Real presence, *Philosophical Topics*, **33**, pp. 235–264.
Pessoa, L. (2008) On the relationship between emotion and cognition, *Nature Reviews Neuroscience*, **9**, pp. 148–158.
Petitmengin, C., *et al.* (2007) Anticipating seizure: Pre-reflective experience at the center of neurophenomenology, *Consciousness and Cognition*, **16**, pp. 746–764.
Piccinini, G. (2010) How to improve on heterophenomenology: The self-measurement of first-person data, *Journal of Consciousness Studies*, **17** (3–4), pp. 84–106.
Protevi, J. (2010) *Larval Subjects, Autonomous Systems, and E. Coli Chemotaxis*, [Online], http://www.protevi.com/john/Deleuze_enaction.pdf
Ruiz-Mirazo, K. & Moreno, A. (2004) Basic autonomy as a fundamental step in the synthesis of life, *Artificial Life*, **10**, pp. 235–259.
Sartre, J.-P. (2004) *The Imaginary: A Phenomenological Psychology of the Imagination*, Webber, J. (trans.), London and New York: Routledge Press.
Shapiro, L. (2004) *The Mind Incarnate*, Cambridge, MA: MIT Press/A Bradford Book.
Sober, E. (1984) *The Nature of Selection*, Cambridge, MA: MIT Press/A Bradford Book.
Sokolowski, R. (2000) *An Introduction to Phenomenology*, Cambridge: Cambridge University Press.
Thompson, E. (1992) Novel colours, *Philosophical Studies*, **68**, pp. 321–349.
Thompson, E. (1995) *Colour Vision: A Study in Cognitive Science and the Philosophy of Perception*, London: Routledge Press.
Thompson, E., *et al.* (1992) Ways of coloring: Comparative color vision as a case study for cognitive science, *Behavioral and Brain Sciences*, **15**, pp. 1–26.
Thompson, E. & Stapleton, M. (2009) Making sense of sense-making: Reflections on enactive and extended mind theories, *Topoi*, **28**, pp. 23–30.
Varela, F.J., Thompson, E. & Rosch, E. (1991) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.
Weber, B.H. & Depew, D.J. (1995) *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection*, Cambridge, MA: MIT Press.
Weber, B.H. & Depew, D.J. (1996) Natural selection and self-organization, *Biology and Philosophy*, **11**, pp. 33–65.
Welton, D. (2011) Can a top-down phenomenology of intentional consciousness be integrated with a bottom-up phenomenology of biological systems?, *Philosophy Today*, in press.
Zahavi, D. (2005) *Subjectivity and Selfhood: Investigating the First-Person Perspective*, Cambridge, MA: MIT Press/A Bradford Book.