Minds, Things, and Materiality

Michael Wheeler
Department of Philosophy, University of Stirling, UK
m.w.wheeler@stir.ac.uk

1. Taking Material Culture Seriously

In a rich and thought-provoking paper, Lambros Malafouris argues that taking material culture seriously means to be 'systematically concerned with figuring out the causal efficacy of materiality in the enactment and constitution of a cognitive system or operation' (Malafouris 2004, 55). As I understand this view, there are really two intertwined claims to be established. The first is that the things beyond the skin that make up material culture (in other words, the physical objects and artefacts in which cultural networks and systems of human social relations are realized) may be essential to the enactment of, and be partly constitutive of, certain cognitive systems or operations. The consequence of establishing this claim is supposed to be that we have a mandate to recast the boundaries of the mind so as to include, as proper parts of the mind, things located beyond the skin. Thus, in talking about the contribution of the world to cognition, Malafouris (2004, 58) concludes that 'what we have traditionally construed as an active or passive but always clearly separated external stimulus for setting a cognitive mechanism into motion, may be after all a continuous part of the machinery itself; at least *ex hypothesi*'. This is the position that, in philosophical circles, is known increasingly as the *extended mind hypothesis* (Clark & Chalmers 1998; Menary forthcoming). Henceforth I shall refer to this hypothesis as *EM*. A stock example will help bring the idea into view. Rumelhart et al. (1986) note that most of us solve difficult multiplication problems by using 'pen and paper' as an external resource. This environmental prop enables us to transform a difficult cognitive problem into a set of simpler ones, and to temporarily store the results of certain intermediate calculations. For the fan of EM, the distributed combination of this external resource and certain inner psychological processes constitutes a cognitive system in its own right.

The second claim that Malafouris needs to establish is that when things-beyond-the-skin achieve the status of being essential to the enactment of, and partly constitutive of, certain cognitive systems or operations, they often do so in virtue of a kind of causal contribution that is, in some way to be determined, a product of those things' essential materiality, rather than in virtue of some other kind of causal contribution. (I shall later identify the former kind of causal contribution as one that involves *vital materiality* and the latter as one that involves only *implementational materiality*.) One implication of this second claim is that for EM be characterized correctly, it needs to pay more than lip service to what we might call the very materiality of material culture.

It is worth noting that, in the context of Malafouris' interests, the pay-off from adopting EM is plausibly nothing less than a reconfiguration of the intellectual landscape inhabited by the discipline known as *cognitive archaeology*. Consider the question 'What does a Palaeolithic stone tool do for the mind?'. Conventional cognitive archaeology is committed to an 'in-the-head' ontology of mind. This condemns material culture to a life outside of cognition proper, and so the things studied by cognitive archaeology are (roughly) things that minds have made and/or used. The cognitive states and processes concerned are not themselves on show in those things, although certain inferences about the nature of those states and processes, inferences of an unavoidably hazardous nature, might be ventured. However, if past ways of thought were not just *expressed in* material culture but were often partly *constituted by* material culture, as EM implies, then cognitive archaeology gets to study past minds in a rather more direct fashion. In fact, the things studied by cognitive archaeology are literally parts of (no longer functioning) minds. If this is right, then the interdisciplinary collective that is cognitive science is poised to welcome a new member of the team.

So where are we going in this paper? In what follows I shall spell out what I take to be the only plausible reading of what is the canonical statement of EM, and argue that, on this reading, the distinctive EM conclusion is purchased using a currency of what I shall call 'implementational materiality'. I shall also submit evidence that Malafouris would judge such implementational materiality to be an inadequate basis for capturing the full cognitive life of things. This goes some way towards breaking the link between Malafouris' vision of what it is to take material culture seriously and EM. To go the rest of the distance, I shall argue that the enactive aspect of Malafouris' approach – recall that we are concerned with the causal efficacy of materiality in the *enactment* and constitution of a cognitive system or operation – is plausibly in tension with EM. If this is right, then taking material culture seriously in the way that Malafouris urges us to, will actually require us to give up on EM.

Although I have just set things up in terms of a critical response to Malafouris' paper, and although I shall refer back to the argument of that paper as the present investigation unfolds, let me stress at the outset that my real target here is certainly not Malafouris, but rather an increasingly widespread tendency in the region of cognitive-scientific space occupied by EM to run together certain importantly different contemporary styles of thinking about thinking. If Malafouris is guilty of this crime, he is far from the only perpetrator. To explain: The wearers of intellectual labels such as situated cognition, embodied-embedded cognitive science, distributed cognition, enactive cognitive science, and EM are wont to march together against the common enemy of a residual Cartesianism in cognitive science. Now, I'm as anti-Cartesian as the next right-thinking cognitive theorist (see e.g. Wheeler 2005), but the fact remains that even if there is some sense in which all these movements are on the same side, their unity against the shared foe serves to obscure certain crucial differences between them. (For a related recent attempt to distinguish EM from other embodiment-based approaches, see Clark forthcoming). I realise that, from the outside, such differences might look nothing more than local spats on a par with the tension between the People's Front of Judea and the Judean People's Front, whose joint opposition to Roman rule is

over-ridden by their doctrinal differences. (If this reference is a mystery to you, go and watch Monty Python's *Life of Brian* – now!). However, whatever the pros and cons of pursuing inter-doctrinal disputes in revolutionary politics may be, the ignoring of fundamental philosophical differences in cognitive science is a sure-fire recipe for confusion. There must come a time when it is right to recognise and to debate those differences. Indeed, sometimes it's the conflicts between thinkers who are, broadly speaking, on the same side that are the most illuminating. As Reg, the leader of the People's Front of Judea, exclaims, "The only people we hate more than the Romans are the… Judean People's Front."

2. So What Exactly *is* the Extended Mind Hypothesis?

Here is what I take to be the canonical (and thus the default) statement of EM:

> [Under certain conditions, the] organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. All the components in the system play an active causal role, and they jointly govern behavior in the same sort of way that cognition usually does. If we remove the external component the system's behavioral competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head. (Clark & Chalmers 1998, 7)

This introduces what has come to be known (by those of us affiliated to the EM club anyway) as *the parity principle*. In broad terms the parity principle states that if there is equality with respect to governing behaviour, between the causal contribution of certain internal elements and the causal contribution of certain external elements, then there is no good reason to count the internal elements concerned as proper parts of the cognitive system while denying that status to the external elements concerned. Parity of causal contribution mandates parity of cognitive status. Notice that, as stated, EM does *not* claim that there are conditions under which the very idea of any internal-external boundary becomes problematic or misleading. It claims rather that there are conditions under which something which counts as a single cognitive system or as a single cognitive process contains some elements which are internal and some which are external. The cognitive process, and thus the mind, are held to be *extended over* that still-in-place internal-external boundary. Of course, this way of talking presupposes that the internal-external boundary at issue is to be fixed by the limits not of the mind, but by those of the brain, or of the skull, or of the central nervous system, or of the skin – take your pick. Clark and Chalmers are not specific on this issue, and for present purposes we need not be either, since, by the lights of the parity principle, there are entities on the external side of any of the latter four interfaces whose causal contribution to behaviour (so the EM theorist argues) will end up counting as cognitive.

With that clarification on board we can ask the following question: what are the benchmarks by which parity of causal contribution is to be judged? Here is the *wrong* way to answer this question. First we fix the benchmarks for what it is to count as a proper part of a cognitive system by identifying all the details of the

causal contribution made by (say) the brain. Then we look to see if any external elements meet those benchmarks. Why is this the wrong way to go? Because it opens the door to the following style of anti-EM argument: we identify some features of, say, internal memory that are not shared by external memory, and we conclude that since the parity principle is not satisfied, EM is false. Rupert (2004) uses this very strategy as the first part of his memory-oriented critique of EM. Here is his own summary of his own argument:

> I argue that the external portions of extended "memory" states (processes) differ so greatly from internal memories (the process of remembering) that they should be treated as distinct kinds; this quells any temptation to argue for [EM] from brute analogy (*viz. extended cognitive states are like wholly internal ones; therefore, they are of the same explanatory cognitive kind; therefore there are extended cognitive states*). (Rupert 2004, 407, my emphasis)

Rupert proceeds to discuss empirical psychological data which putatively indicate significant differences between (a) the profile of internal memory and (b) the profile of certain external resources, as such external resources figure in the process of remembering. According to Rupert, such differences tell against any attempt to see the latter phenomena as being of the same explanatory kind as the former. For example, there are psychological experiments which show that internal memory is sensitive to what is called the generation effect. Where this effect is in evidence, subjects gain a mnemonic advantage by generating their own meaningful connections between items be learned. Rupert argues that the generation effect will simply not occur in some extended 'memory' systems (e.g., in a system according to which, during recall, the subject refers to a notebook in which the paired associates are accompanied by connection sentences produced by those subjects during learning, but which were entered into the notebook by the experimenter). He concedes that it might occur in others (e.g., in a system according to which, during recall, the subject refers a notebook in which the paired associates to be learned are accompanied by connection sentences produced and entered by the subjects during learning). In the latter case, however, he concludes that the effect is an accidental feature, rather than a defining dimension, of the memory system.

For the sake of argument, let's just agree that Rupert is right about the presence of such differences here. What the fan of EM needs to do is simply refuse to accept that one should allow the extant details of internal memory to set the benchmarks for what counts as memory in general. To be clear, the Rupert-style argument under consideration isn't suspect in virtue of being anti-EM. Rather it's suspect because it begs the question against EM by assuming that what counts as cognitive should be fixed by the fine-grained profile of the inner. Such question-begging can be avoided, and Rupert's criticism resisted, if we adopt the following alternative strategy for saying what the benchmarks are by which parity of causal contribution is to be judged. First we give an account of what it is to be a proper part of a cognitive system that is essentially independent of where a candidate element happens to be located with respect to the internal-external boundary (however that boundary is to be determined). Then we look to see where cognition falls – in the brain, in the non-neural body, in the environment, or, as the EM

theorist predicts will sometimes be the case, in a system that extends across all of these aspects of the world.

Rupert sees this sort of response coming, and so develops his memory-oriented critique further by arguing that any attempt to fix a generic kind that would subsume internal and extended systems would need to be so devoid of detail (in order to subsume all the different profiles) that it would fail to earn its explanatory keep. But this seems wrong. Indeed, it's important to note that we would surely not intuitively withdraw the epithet 'memory' from an internally located system which did not exhibit the generation effect, but which continued to achieve (something like) the selective storage and context-sensitive retrieval of information, so why should we withdraw that epithet from an extended system with a similar profile? But if that's right, why think that exhibiting the generation effect is a defining dimension of memory, rather than an accidental feature? This gives us some reason to think that there must be a generic account of what memory is that covers both cases, and that has explanatory bite.

Here it is worth pausing momentarily to consider a different response to the clearly prejudiced fix-according-to-the-inner strategy. According to this response, we should first work out the details of what the brain does, and then remove from our list of features any details that are inessential to that contribution *as cognitive*. That way we will be able to rule out the arbitrary exclusion of external elements and arrive at a viable set of benchmarks for what it is for something to count as a proper part of a cognitive system. However, any *decent* version of this response must, it seems, collapse into a version of the strategy that I am recommending, since in order for some detail of a causal contribution to be judged inessential to that contribution as cognitive, one must have access to an independent theory of the cognitive. The alternative, which would involve ruling out a detail purely on the grounds that it is not shared by some external element under consideration would, of course, beg the question against the opponent of EM.

So what sort of overarching theory of the cognitive is favoured by EM theorists? As Clark (forthcoming) notes, the fact is that '[a]rguments in favour of [EM] appeal mainly, if not exclusively, to the *computational role* played by certain kinds of non-neural events and processes in online problem-solving'. In other words, EM theorists overwhelmingly conceive of cognition as a matter of *information processing*. Their distinctive observation is that, given this view of what cognition is, extra-neural factors – including the stuff of material culture – may, in some cases anyway, realise the target phenomenon just as readily as neural tissue. For example, taking it that memory is, at least in part, a matter of the selective storage and context-sensitive retrieval of information, the EM theorist with a cognitive-archaeological bent might contend that information that is poised appropriately for context-sensitive retrieval may be stored in a Mycenaean Linear B tablet just a readily as in a Mycenaean brain.

3. Extended Functionalism and Implementational Materiality

Of course, not any old kind of information processing profile will do here. To say that it would would be to fall prey to Rupert's worry about explanatory inefficacy. No, genuine cognition will be found only in a (perhaps rather small) subset of information processing systems. But whatever the detailed story one eventually might arrive at via painstaking philosophical and empirical research, the basic commitment to cognition as a kind of information processing is precisely what Malafouris thinks of as the kind of residual cognitivism that stymies progress towards understanding the rich way in which things may have cognitive lives. As Malafouris (2004, 55) puts it, 'material culture has a place in the mind only as a disembodied digit of information written somehow on the neural tissue'. But, he argues, material culture has a far richer cognitive life than that image allows. Using striking examples from the contributions of artefacts within Mycenaean funeral rituals, he suggests that the role of 'intentional or non-intentional mnemo-technical artefacts and practices is far more dynamic and dialogical' and so artefact-involving memory is reveals itself to be a 'process of active discovery spanning the monumental and the minute, the conspicuous and the commonplace, iconicity and iconoclasm' (Malafouris 2004, 57).

What is going on here? Although Malafouris doesn't put things in quite the way that I'm about to, his view, I think, is that those objects whose contribution to cognitive life is merely a matter of enabling enhanced performance in the storage and manipulation of information display only what we might call *implementational materiality*. In other words, given that what is important in understanding the mental status of such objects is that they perform certain computational functions that are allegedly distinctive of cognitive processing, the materiality of those objects is relevant only as the explanation of how the computational functions in question are *implemented in the physical world*. The real explanatory action is reserved for the more abstract, and in some sense non-material (that is, disembodied), information processing story.

But is EM's standard reliance on the concept of information the real issue? I don't think so. Let's allow the EM theorist to shed the grand theory of cognition as information processing. Where might she go next? Well, she might preserve much of what is important about the computational approach without actually thinking of cognition as computation. In other words, she might declare herself to be a *non-computational functionalist*. According to the traditional formulation of functionalism in the philosophy of mind, the canonical statement of which is arguably due to Putnam (1967), a mental state is constituted by the causal relations that it bears to sensory inputs, behavioural outputs, and other mental states. The information processing theory of cognition, which glosses functional role as computational role, is just one possible version of functionalism, so conceived. As an alternative to computational functionalism one might, for example, gloss one kind of non-computational functional role in terms of maintaining systemic stability in a homeostatic dynamical system by keeping the values of certain critical variables within certain limits. What this tells us is that the standard way of unpacking EM in information processing terms is just one possible version of a position that Clark has dubbed *extended functionalism* (Clark forthcoming). According to extended functionalism, a mental state is constituted by the causal relations that it bears to systemic inputs, systemic outputs, and other systemic states. (This is merely a more general formulation of the functionalist line, one that

in principle allows the borders of the cognitive system to fall somewhere other than the sensory-motor interface of the organic body.) The very possibility of a non-computational extended functionalism shows that information-processing (at least as construed in standard cognitive-scientific terms) cannot be at the centre of the present issue. For all forms of extended functionalism will be just as committed to implementational materiality as the more narrowly defined information-processing variety.

To illustrate this point, consider the following example. Air traffic controllers typically co-ordinate their activity using flight strips – bands of paper printed with flight information (e.g., airline name, flight number, and type of aircraft, plus the speed, level, and route of the flight plan, both as requested and as authorized). When thinking about how air traffic controllers succeed in their complex, high-pressure job, and about how contemporary technology may enhance that success, one is inclined to focus, naturally enough, on the information carried by these strips. But this is not the only contribution of the strips. Mackay et al. (1998) argue convincingly that the physical embodiment of the strips supports a number of workplace strategies employed successfully by the controllers. For example, individuals often hold the strips as reminders to perform some action, or slide them to the left or right to indicate certain conditions (e.g., two planes in a potential conflict situation). Moreover, two controllers may work simultaneously on the same strip-holding board, using body language to signal the importance of particular movements or rearrangements of the strips.

From a practical perspective, this recognition of the non-informational contribution of the flight strips is far from idle. The testimonial evidence suggests that a number of previous attempts to introduce new computer technology into air traffic control may ultimately have been rejected as unworkable by the controllers precisely because the proposed replacement systems attempted to reproduce the straightforwardly informational aspects of the flight strips while ignoring the extra factors. Thus, and recognizing that the non-informational interactions supported by the flight strips would be difficult to reconstruct in any keyboard/monitor interface, Mackay et al. advocate the use of augmented electronic strips. From a theoretical perspective, and pending further analysis, one might at least consider the first two of the identified contributions of the flight strips (memory, the modelling of possible states of affairs) to put those artefacts in the ballpark to be considered elements in the cognitive architecture of the flight controller. And now note that nothing about this story undermines the extended functionalist line. The flight strips reveal their (provisionally) cognitive status precisely by carrying out certain functionally defined systemic roles that are distinctive of cognition (memory, the modelling of possible states of affairs). The materiality of the flight strips is thus implementational in character.

It is easy enough to see through to theoretical heart of this issue. In order to fly, EM needs to embrace a key feature supported by functionalist theorizing – namely *multiple realizability*. A little philosophical history will help here. Functionalism (in its non-extended form) freed physicalist philosophy of mind from a kind of neural chauvinism. If our mental states were constituted by their functional roles, and the material contribution of our brains was merely implementational in character, then robots, Martians, Klingons, and gaseous

creatures from the outer limits of the universe could all join us in having mental states, just so long as the physical stuff out of which they were made could implement the right functional profiles. Stretching the word 'skin' to include boundaries made of tin and gas, traditional functionalism bequeathed to the mind what we might call within-the-skin multiple realizability. And within-the-skin multiple realizability requires within-the-skin implementational materiality. But now extended functionalism merely plays out the same logic beyond the skin. If the specific materiality of the substrate doesn't matter to cognition, outside of the fact that it must be able to support the required functional profile, then what, in principle, is there to stop things-beyond-the-skin counting as proper parts of a cognitive architecture? Nothing, that's what. And this beyond-the-skin species of multiple realizability, which is just another way of characterizing the core philosophical commitment of EM, requires beyond-the-skin implementational materiality. If we look at things this way, the really radical and revolutionary movement was functionalism, not EM. EM simply makes manifest one of the implications of functionalism. In other words, EM is just a footnote to Putnam

4. Enactivism and Vital Materiality

One could, I suppose, develop a version of EM (as standardly conceived) without buying into functionalism, *if* one could have multiple realizability without functionalism. Churchland (2005) has argued recently that the latter is possible. But even if one could have EM without functionalism, one couldn't have EM (as standardly conceived) without multiple realizability, and so one couldn't have EM (as standardly conceived) without implementational materiality. Given that Malafouris finds implementational materiality wanting as the basis of a full account of the cognitive life of material culture, this goes some way towards breaking the link between EM and what he calls 'taking material culture seriously'. Now recall that taking material culture seriously means to be 'systematically concerned with figuring out the causal efficacy of materiality in the enactment and constitution of a cognitive system or operation' (Malafouris 2004, 55). So what is this other species of material causal efficacy, the one that, according to Malafouris, goes beyond mere implementational materiality, and which matters to the cognitive life of things? Let's label this form of material causal efficacy 'vital materiality'. What, then, is vital materiality? To answer this question we need to turn to Malafouris' compelling image of the potter at her wheel.

> [The] cognitive map of knowledge and memory may well be extended and distributed in the neurons of the potter's brain, the muscles of the potter's body, the 'affordances'… of the potter's wheel, the material properties of the clay, the morphological and typological prototypes of existing vessels as well as the general social context in which the activity occurs. (Malafouris 2004, 59)

So far, nothing new seems to be on the table – or rather the wheel. Indeed, it looks as if what we have here is EM. It might seem, therefore, that the lesson of the potter's wheel is that we should replace our view of cognition as residing inside the potter's head, with that of cognition as spatially distributed over brain, body and world. But Malafouris develops his case in a different direction:

It is at the potter's fingers that the form and shape of the vessel is perceived as it gradually emerges in the interactive tension between the centrifugal force and the texture of the wet clay. Materiality enters the cognitive equation at a much more basic level, shaping the phenomenology of what Searle has defined as the 'Background' i.e., the set of non-representational mental capacities that enable all representing to take place… In other words… we should replace our view of cognition as residing inside the potter's head, with that of cognition enacted *at the potter's wheel*. (Malafouris 2004, 59)

Clearly, we need to understand the term 'enacted'. As the references given by Malafouris (2004, 57) indicate, the relevant interpretation of this idea is provided by (i) the enactive understanding of cognition (henceforth *enactivism*), as developed and defended by Varela et al. (1991; see also Thompson forthcoming), construed as a development of (ii) the autopoietic theory of Maturana and Varela (1980). And there's the rub. For as far as I can see, on that interpretation of the pivotal idea, cognition enacted *cannot* be cognition extended. Here's why.

The smart money says that the theory of autopoiesis (an all-too-brief introduction to which will be given in a moment) is a non-negotiable component of enactivism. This might come as a surprise to some fans of the enactivist 'bible' *The Embodied Mind* (Varela et al. 1991) since that text doesn't foreground autopoiesis as a term, but the fact is that many of the component concepts from autopoietic theory, such as autonomy and structural coupling, all appear there in pivotal roles, and are used in ways identifiable from the theory of autopoiesis. More recently Thompson, in what I believe is poised to become the leading post-*Embodied-Mind* development of enactivism, explicitly treats the theory of autopoiesis as one of the conceptual keystones of the position (Thompson forthcoming, chapter 5). This suggests that getting straight about autopoiesis will help us to understand the enactivist's core commitments. The bad news for the uninitiated is that the theory of autopoiesis is so devilishly complex that I can't hope to do it justice in the space available here. The good news is that I don't need to worry about this exegetical shortfall since I'll be concerned with the overall shape of things, not the fine-grained details. Under such circumstances, the roughest of rough guides will do.

Autopoietic systems form a subset of *self-organizing* systems (Weber and Varela 2002; Di Paolo 2005), where a self-organizing system is one in which the intra-systemic components, on the basis of purely local rules of interaction, interact with each other in nonlinear ways to produce the emergence and maintenance of structured global order. More specifically, autopoietic systems form a subset of *autonomous* self-organizing systems, where an autonomous system is one in which the constituent processes '(i) recursively depend on each other for their generation and their realization as a network, and (ii) constitute the system as a unity in whatever domain they exist' (Thompson forthcoming, 64). Examples of autonomous self-organizing systems include cells, nervous systems and insect colonies. However, while self-organization and autonomy are necessary for autopoiesis, they are not sufficient. To be autopoietic, a self-organizing autonomous system must also, through its own endogenous activity, produce and maintain a *physical* boundary that distinguishes that system as a material unity (a

minimal self) in the space in which it exists. The paradigm example of an autopoietic system is the single cell (a network of chemical reactions that produces its own membrane), and one might reasonably think of the concept of autopoiesis as capturing the distinctive mode of systemic organization that is realized in material living systems as metabolism (Di Paolo 2005). Crucially, the self-distinguishing process characteristic of autopoiesis takes place in the face of physical perturbations from the system's environment. This is what autopoietic theorists call *structural coupling*, a process in which an autopoietic system encounters its environment while maintaining its organization.

Now for the point of all this. According to Maturana and Varela, autopoiesis is necessary and sufficient for life (see e.g. Maturana and Varela 1980, 82). In other words, for Maturana and Varela, any living system is an autopoietic system and any autopoietic system is a living system. But what about cognition? One striking claim in the core literature is that living 'simply' *is* cognition: '[l]iving systems are cognitive systems, and living as a process is a process of cognition' (Maturana 1970, 13). How could this be? For Maturana, cognition is effective (i.e. viability-maintaining) activity in a domain of interactions defined by the autopoietic system's organization. In this context, then, *enaction* is the process by which significance or relevance is brought forth through the viable structural coupling of the autopoietic system with its environment. The idea here is that the autopoietic organization, in establishing the distinction between the self-maintenance and the collapse of the system as a unity, institutes a norm of survival, and thereby the significance or relevance of certain environmental perturbations as either leaving the system organizationally intact or resulting in its disintegration. (Notice that the tight link between cognition and autopoiesis plausibly introduces something that we might be moved to call 'vital materiality' in the case of the former, given the stress in the latter on the endogenous construction and maintenance of a self-distinguishing *material* boundary. However, as I shall explain later, I don't think this is the idea that Malafouris has in mind.)

Of course, it is possible to hear Maturana's striking claim in more than one way – as advancing either (a) the view that life and cognition are identical, or (b) the view that living systems are a subset of cognitive systems (Thompson 2004). For the moment let's work with (a). I'll come back to (b) later. If the living system is identical with the cognitive system, then the boundary of the living system will coincide with the boundary of the cognitive system. And it's this that (finally) generates the inconsistency with EM, given, that is, that we accept what I take to be a *highly* plausible claim, namely that where we have an extended cognitive system, the living system (the organism, although not necessarily the phenotype) remains bounded by its skin. The external elements that, according to EM, count as proper parts of the cognitive system do not thereby become proper parts of the living organism. They remain abiotic. In other words, and pace Scott Turner (2000), organisms don't extend even if minds and, as Dawkins (1982) pointed out, phenotypes do. But if, in cases of extended cognition, the boundary of the cognitive system and the boundary of the living system come apart, as I have suggested, then one cannot simultaneously sanction EM and identify the cognitive system with the living system. Since autopoietic theory (and thus enactivism), on our current interpretation, is committed to identifying the cognitive system with

the living system, it simply cannot endorse EM. In other words, enacted minds are not extended minds.

So far so bad for the attempt to bring EM and enactivism together. But the game is not yet over, because a more complex understanding of the relationship between life and cognition has recently entered the autopoiesis-enactivist literature (Di Paolo 2005; see also Thompson forthcoming). In his later work, Varela (e.g. 1991) argued that cognition is a process of (what he dubbed) *sense-making* and that living is just such a process. Sense-making enriches the autopoietic picture of an enacted domain of significance by introducing the idea that that domain is 'a place of *valence*, of attraction and repulsion, approach or escape' (Thompson 2004, 386). Varela (1991) and Thompson (2004) both suggest that autopoiesis is sufficient to instantiate sense-making (hence living is a process of sense-making). However, Di Paolo (2005) has argued – convincingly as Thompson (forthcoming) notes – that sense-making requires something over and above what we might now think of as *raw* autopoiesis. This is because raw autopoiesis bestows only a kind of robustness or conservation of systemic organization in the face of environmental perturbation. The system doesn't alter its behaviour in response to changes in its environment. It either survives any perturbations it experiences, or it doesn't. As Di Paolo usefully puts it, the norm of survival as established by raw autopoiesis is an all-or-nothing affair. But sense-making (as Thompson's talk of attraction, repulsion, approach and escape indicates) requires a system to be sensitive to graded differences between states. The organism needs to monitor *how it is doing* with respect to the norm of survival and to regulate its behaviour accordingly in order to improve its situation. In other words, it needs to be an *adaptive* system.

Crucially, while adaptivity is a phenomenon over and above raw autopoiesis, it needs to be established on the basis of autopoietic principles. Without the connection to the self-distinguishing process of autopoiesis, the meaning generated by adaptivity and sense-making would not be established as original to the activity of the system (meaning for the system), but would merely be attributable to the system's activity by some external observer (meaning merely ascribed to the system) (Di Paolo 2005; Thompson forthcoming, 231). So, given that raw autopoiesis is necessary but not sufficient for sense-making, and given that sense-making is to be identified with cognition, raw autopoiesis is necessary but not sufficient for sense-making, which means that cognition is sufficient for raw autopoiesis. But now since on our revised picture raw autopoiesis remains necessary and sufficient for life – Di Paolo never suggests that a raw autopoietic system wouldn't be alive, only that it wouldn't be a cognitive system – being a cognitive system is sufficient for being a living system. And that means that the tension between enactivism and EM is still in force. If the enactivist did endorse EM, she would (among other things) be claiming (a) that an extended cognitive system is an autopoietic system, and thus (b) that an extended cognitive system is itself (and not merely contains) a living system. As far as I can see (a) is debatable, but in any case (b) violates our highly plausible thought concerning organismic extension (or rather the lack of it). Since the enactivist cannot give up on the co-location of the cognitive system and the living system, she cannot endorse EM. To repeat, enacted minds are not extended minds.

I conclude that there is a prima facie case for the incompatibility of EM and enactivism. There are, of course, further moves that could be made. For example, one might try to unpack a recognizably and genuinely enactivist notion of cognition in which there is no (explicit or implicit) *necessary* dependence on the phenomenon of autopoiesis, thereby creating conceptual space for the previously mentioned view that living systems form a subset of cognitive systems. On this account, being an autopoietic (living) system would not be necessary for a system to be cognitive. That would allow there to be a cognitive system that is not itself identical with some living system, and that would establish consistency with EM, since the cognitive system could be extended when the living system is not. The plausibility of such an approach cannot be ruled out in advance. At the very least, however, I hope to have gestured in the direction of the sorts of sacrifices and challenges that the project of developing such a view would face, given the autopoietic heritage of enactivism, as standardly conceived.

Aside from (what I have argued) amounts to a parting of the embodied-embedded ways (and note that I've said nothing about which of our two views might be correct), it's also remains unclear, from what I've said about enactivism, exactly why that approach might be the road to imbuing material culture with something that Malafouris would welcome as vital materiality. I shall finish with a sketchy remark on this point. The key here is in Malafouris' claim that the material nature of the interactions between the potter's fingers, the wheel, and the clay means that materiality shapes the fundamental phenomenological structure of the situation. There is no doubt that this is broadly in line with enactivist theorizing, which holds experience to be bodily in character. Let's say that this is right, and that in the end it's the way that the materiality of things, alongside our own embodiment, structures our fundamental phenomenological space, that gives us a concept of the vital materiality of material culture. Without further argument, it is at least uncertain that this phenomenological point *alone* could ever mandate the claim that things-beyond-the-skin *must* be counted parts of the cognitive system, as opposed to mind-external elements that play a causal role alongside mind-internal elements in structuring experience. More, it seems, remains to be said.

References

Churchland, P. M., 2005. Functionalism at forty: a critical retrospective. *Journal of Philosophy* 102:1, 33-50.

Clark, A., Forthcoming. Pressing the flesh: exploring a tension in the study of the embodied, embedded mind. *Philosophy and Phenomenological Research.*

Clark, A. & Chalmers, D., 1998. The extended mind. *Analysis* 58 (1), 7-19.

Dawkins, R. 1982. *The Extended Phenotype: the Long Reach of the Gene.* Oxford: Oxford University Press.

Di Paolo, E.A., 2005. Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences* 4:4, 429-52.

Mackay, W. E., Fayard A.-L., Frobert, L. & Medini, L., 1998. Reinventing the familiar: exploring an augmented reality design space for air traffic control, in *Conference Proceedings on Human Factors in Computing Systems (CHI 1998)*. New York: ACM Press/Addison-Wesley, 558-65.

Malafouris, L., 2004. The cognitive basis of material engagement: where brain, body and culture conflate, in *Rethinking Materiality: the Engagement of Mind with the Material World*, eds. E. DeMarrais, C. Gosden & C. Renfrew. Cambridge: McDonald Institute for Archaeological Research, 53-61.

Maturana, H. R., 1970. Biology of cognition, in (Maturana and Varela 1980), 2-58.

Maturana, H. R. & Varela, F. J., 1980. *Autopoiesis and Cognition: the Realization of the Living*. Dordrecht: Boston Studies in the Philosophy of Science, Vol. 43, D. Reidel.

Menary, R. (ed.), Forthcoming. *The Extended Mind*. Aldershot: Ashgate.

Putnam, H., 1967. Psychological predicates, in *Art, Mind and Religion*, eds. W.H. Capitan & D.D. Merrill. Pittsburgh: University of Pittsburgh Press.

Rumelhart, D.E., Smolensky, P., McClelland, J.L. and Hinton, G., 1986. Schemata and sequential thought processes in PDP models, in *Parallel Distributed Processing: Explorations In The Microstructure Of Cognition, Vol. 2: Psychological And Biological* Models, eds. J.L. McClelland and D. Rumelhart. Cambridge, Mass.: MIT Press, 7-57.

Rupert, R., 2004. Challenges to the hypothesis of extended cognition. *Journal of Philosophy* 101: 8, 389-428.

Scott Turner, J. 2000. *The Extended Organism: the Physiology of Animal-Built Structures*. Cambridge, Mass.: Harvard University Press.

Thompson, E., 2004. Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences* 3, 381-98.

Thompson, E., Forthcoming. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, Mass.: Harvard University Press. The page numbers given in quotations are taken from a pre-publication manuscript.

Varela, F. J., 1991. Organism: a meshwork of selfless selves, in *Organism and the Origin of Self*, ed. A. Tauber. Dordrecht: Kluwer Academic Publishers, 79-107.

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

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

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