

# Agency is Distinct from Autonomy

**Fred Cummins**

UCD School of Computer Science and Informatics  
University College Dublin  
*fred.cummins@ucd.ie*

Received December 2013; accepted June 2014; published Autumn 2014.

## Abstract

Both autonomy and agency play central roles in the emerging enactive vocabulary. Although some treat these concepts as practically synonymous, others have sought to be more explicit about the conditions required for agency over and above autonomy. I attempt to be self-conscious about the role of the observer (or scientist) in such discussions, and emphasise that the concept of agency, in particular, is deeply entwined with the nature of the observer and the framing of the observation. This is probably well known to enactivists, but runs the risk of being badly misunderstood if it is not made explicit. A heightened awareness of the role of the observer in the attribution of agency may allow us to make advances in questions in which progress is hindered by assuming a single split between subject and object. I argue that human experience is characterized by our embedding in webs of meaning arising from our participation in systems of many sorts, and that this richness demands a corresponding lightness of touch with respect to the identification of agentic subjects.

**Keywords:** agency; autonomy; enactivism; cognition; mind.

## 1. Introduction

The treatment of agency in human intercourse is a matter of the highest importance. It underpins the enlightenment notion of human rights. It informs, or should inform, the manner in which society apportions responsibility for actions, including those boundaries that delimit the criminal and the insane. The attribution of agency is one of the principal themes that informs and defines all religious traditions. When the enactive and mind-and-life schools choose to make agency a foundational concept, they must do so with a sense of gravitas, for the issues at stake go beyond novelty in the cognitive and behavioral sciences or philosophy of mind. In what follows, I will consider the entanglement of the twin concepts of autonomy and agency as they feature in

the enactive literature. At stake is whether these terms pick out distinct concepts, and more importantly, whether agency can and should be naturalised, and what that would mean.

I begin by acknowledging the consequences of a constructivist approach for the practice of science, which demands that we recognize the social and historical context in which the attribution of agency has been restricted to a single kind of split between agentive subject and associated world. Some careful attempts to characterize agency over and above autonomy are discussed in detail, leading to the cautious stance that an overly rigid attribution of agency to this or that system may blind us to the important distinction between the domain of the system and the domain of the description of the system. This caution allows us to recognize the plurality of systems among which our lives unfold, that provide many and varied examples of systems that may appear agentive. The hope is that this caution will allow us to recognize that in attributing agency here, or there, we are at the same time identifying our own being with diverse forms of systematic organisation and hence with diverse systems of values that thereby arise.

The enactive agenda necessarily has one foot in scientific practice and another in the philosophy of science (there may be more than 2 legs). As Stewart noted “the paradigm of enaction is ontologically nonobjectivist—or to put it more positively, radically constructivist” (Stewart et al. 2010). Constructivist agendas have not found easy integration into mainstream scientific practice, and indeed many, perhaps most, areas of scientific inquiry get by just fine with the presumption that they are uncovering the structures and form of a mind-independent world. For those of a constructivist bent, this is, of course, no longer tenable for “sciences of the mind”, or, by extension, for any of the Human Sciences.

A necessary consequence of adopting a constructivist perspective is that explanation is necessarily bounded. In the spirit of pragmatic explanation broadly construed, the practice of explanation, description and modeling is not to fix this or that phenomenon within a single static account of the world, but to develop an understanding that is appropriate to a specific domain, and that may serve in a broader array of accounts, some more specific, some more general, but none aspiring to be ground truth. The pragmatic boundedness of explanation here is neither the logical kind espoused by Pierce (Murphy and Rorty 1990) nor the unbridled categorical cornucopia of Ryle (Ryle 1949), but demands that explanation be couched in terms appropriate to specific domains. The biological concept of autopoiesis identifies a particularly important form of autonomy that is defined with respect to the chemical domain (Maturana and Varela 1991). A more general concept of autonomy, in Varela’s definition, requires *inter alia* that the processes that characterize an autonomous system be a unity “recognizable in the space (domain) in which the pro-

cesses exist” (Varela 1979: 55). In an informative example, Varela suggests that “the accidental collision of two running animals, as a bodily encounter of living systems, is not a biological phenomenon (even though it may have biological consequences), but the bodily contact of two animals in courtship is.” (Varela 1979: 42).

Most scientific practice is not done within a constructivist framework, and an appreciation of the limited nature of explanation, of the finiteness of domains of discourse, is not something that can be taken for granted among scientists generally, and still less so in the conversation between science and society more generally, where science is still almost universally regarded as the source of certainty, and the ultimate arbiter of disputes. As the enactive agenda develops, and seeks to provide a set of concepts applicable in domains from the biochemical to the social, it must exercise caution if it is not to be enthusiastically misunderstood as a means of establishing certainty with respect to many vexed notions that it takes as central. Among these are the twin terms of autonomy and agency. This is a first challenge we face.

In its principled rejection of the Cartesian split of things into {inner, mental, and subjective}, versus {outer, material, and objective} realities, enactive theory finds itself in the lexically challenged business of talking about human experience and behavior without being able to lean confidently upon any psychological predicates whatsoever. Yet we cannot do without these. To address this, it is necessary to recognize a second consequence of a radical constructivist approach to understanding: As we learn about phenomena in various domains, so too we are learning about that which we are.

Maxim: Along with asking “what is this thing that we see”, we need to ask “what are we that we should see such? ”.

A corollary of this is that our understanding of *all* psychological predicates should be taken as tentative, and subject to change.

The emphasis in that maxim is on the “we”. The sociocultural background in which scientific practice has developed has the striking characteristic of attributing agency in human affairs exclusively to the individual person, conceived of as a discrete organic unit. Democratic societies that emerged after the Age of Revolutions are founded upon the notion of individual agency. The prevalent Christian theology, and post-Reformation Protestant ideology position the concepts of culpability and responsibility squarely within the individual. It is against this backdrop that the science and practice of psychology emerged in the latter half of the Nineteenth Century, and the unit of psychology, even social psychology, is the individual person, conceived of as a singular mind housed within a singular body.

The emerging enactivist position has the potential to move our discourse beyond the limitations of Protestant theology and 19th Century psychology. The language that is emerging, and being carefully tended and curated, allows explanation to countenance multiple perspectives, even mutually incompatible perspectives. In its parallel treatment of system-internal and system-external perspectives, it allows for a richer, plural, understanding of the interactions among complex systems such as cells, persons, or social groups, and a recognition of disparate value sets that influence their evolution and interaction. But a positivist view of scientific understanding and a restriction of agency to single human individuals are both deeply entrenched in our ambient belief systems and practices. Careful exegesis of the conceptual fulcrums of the enactive approach is indispensable, and the treatment of agency in particular will be of singular importance. It would be unwise to be dogmatic here.

## 2. Pinning Agency to the Wall

Autonomy and agency are absolutely central to the enactive approach.<sup>39</sup> The somewhat vague notion of autonomy was given a more precise definition by Varela as “organizational closure” (Varela 1979: 58). On this view, autonomous organization is characterized by circular closure among a suite of processes that collectively constitute a persistent dynamic identity that engages in regulated exchanges with its surround. The most discussed example, by a substantial margin, is the caricature of a single bacterium ascending a chemical gradient through chemotaxis. This example, trotted out repeatedly with differing degrees of attention to biochemical detail, serves as the exemplary embodiment of the autonomous and agentic system. As an illustrative case, it allows discussion of value, as the nutrient is unambiguously a “good thing” from the point of view of the bacterium itself. It allows discussion of sense-making by linking the effective coping of the bacterium in a variable environment with its own metabolic requirements. It illustrates the contrast between a perspective anchored to a specific system, and a view from nowhere. In short, the chemotactic bacterium serves as the central myth of the enactive approach, in the sense of a narrative that serves to structure many discussions that need to be had as we apply systems thinking to our own selves.

But real bacteria are vastly different, more complex, more tightly embedded in their environments, and more social than this. The bacterium of this oft-repeated illustration is described as a minimal *mechanism*, requiring a single sensor capable of detecting ambient glucose concentration, a means of locomotion with a directed mode and an undirected mode, and a probabilistic link between the slope of the ambient gradient and the likelihood of switching

---

<sup>39</sup> I do not make any strong distinction between enactive theory and mind and life positions within philosophy. Both terms cover a range of positions, with very substantial overlap.

between the two modes. Viewed as a mechanism, it is indistinguishable from any machine concocted by a designer, and indeed, as a character in our narrative, it could be argued that it has indeed been designed. There is nothing to be found in the mechanism that warrants the attribution of agency. That attribution comes instead from our understanding, as scientists, that chemotactic locomotion can be understood as serving the continued existence of the cell as an organisationally closed set of processes. A functional description, by us as scientists, is predicated upon our passing adoption of the perspective of the bacterium, which is the domain for whom the ambient glucose is meaningful, thereby licensing talk of “function”.

In discussion of the caricature of the bacterium, two related issues seem to become entangled. One issue concerns the difficult task of differentiating between machines that are designed, and machines that have evolved. The second lies in distinguishing between “merely” autonomous phenomena such as tornados and flames on the one hand and apparently agentic (and autonomous) cells on the other. With respect to the first distinction among machines, there is no overt marker of the locus of the designer, or the origin of the organising principle that characterises the machine. Thus that distinction is made on the basis of our knowledge of the history of the machine. With respect to the second, there is much discussion that remains to be had, but the issues at stake are importantly different.

An informative example of the confusion of the concepts of autonomy and agency is provided by two related papers. In Rohde and Stewart (2008), it is argued that *autonomy* is a complex property, similar in some respects to such complex notions as intentionality or intelligence. When faced with a system (natural or artificial), the ascription of autonomy thereto can not be based on a set of necessary or sufficient conditions. But the authors are not happy with the “as if” ascription that falls out of adopting what Dennett has called the “intentional stance” with respect to the system (Dennett 1989). Taking a constructivist perspective, the authors seek to improve on mere ascription through appeal to generative mechanisms that can give rise to the phenomenon of autonomy. Taking a cue from that suggestion, then, Barandiaran et al. (2009) attempt to define the concept of *agency*, first through mere description, and then improving and deepening that characterization through the proposal of a set of generative mechanisms that can give rise to agency. The switch from a generative description of autonomy to a generative description of agency provides a telling example of the manner in which these two concepts have become entangled. Note that the explanatory move common to both these papers is independent of the distinction I am drawing here. Both papers seek to ground an account of a complex phenomenon in something more than mere description, and both do so by appealing to generative mechanisms. Whether this succeeds or not is not at issue here. What is at issue is

whether autonomy and agency are separate dissociable concepts. Let us look at some examples from the recent literature:

Moreno and Etxeberria (2005) contrast the "actions" performed by a stone in a river that keeps water from flowing, and by bacteria in milk that ferment it. Only the latter, they say, can lay claim to agentive goings on, and they ground this claim in the role of the fermentation from the perspective of the bacterium. This contrast seems particularly clear, but mainly because the rock is not "doing" anything. Absent activity on the part of the agent, agency is invisible. Barandiaran et al. (2009) consider a suite of more interesting and challenging contrasts that provide guidance in the application of three criteria: *Individuality* (is the system an individual in the above sense?), *Interactional Asymmetry* (is the system the active source of interaction with its environment?), and *Normativity* (is the norm that shapes the interaction generated by the system itself?). Simultaneous satisfaction of these three conditions serve to pick out the chemotactically locomoting cell as the only instance of full agency, while several other carefully chosen candidates fail to tick at least one of the three boxes. Individuality corresponds roughly to the definition of autonomy as organisational closure, as discussed above. Both cells and tornados might reasonably meet the criterion of systemic individuality. We therefore need to consider whether the other two conditions succeed in going beyond autonomy and to pick out agency as a distinct and definite concept.

The requirement of normativity refers to the interpretation of a behaviour as subserving goals, and Barandiaran et al. (2009) wish to restrict the attribution of agency to systems which generate their own goals. Many authors have recognized that agentive behaviour occurs in the service of goals, indeed its goal directed nature may be the very feature that allows a continuous stream of movement to be parsed by an observer into discrete behaviours for consideration in the first place. Some have opined that goals alone are sufficient for agency (Beer 1995), but most researchers, especially those who seek to distinguish the agency of the living from the functional carry-on of devices and robots, have insisted that agency requires that the goals arise from the needs and identity of the system itself (Weber and Varela 2002; Maes 1993; Christensen and Hooker 2000; Kauffman 2002; Deacon 2011, and others). Upon this rock, many have foundered, even Immanuel Kant himself, for whom the intrinsic purposes of the living posed an insurmountable challenge within a Newtonian metaphysical view. Unlike Kant, we have available to us a richer metaphysical armoury and the powerful set of concepts that stem from the description and study of complex systems (Weber and Varela 2002). However, even with this endowment, there is not available to us any litmus-test for distinguishing between intrinsically generated and extrinsically imposed goals in the shaping of behaviour.

A first example adduced by Barandiaran et al. (2002) meets the criterion of intrinsic normativity, but is still regarded as non-agentive. Consider a mother cat and several kittens. One of the kittens is moved by the mother closer to her body to keep it warm. The kitten meets the criterion of dynamic identity, and it also satisfies the intrinsic normativity criterion because “the system-environment coupling is satisfying the norm of keeping the kitten’s temperature within viability boundaries” (Barandiaran et al. 2009: 5). Agency is lacking, the authors contend, because the movement has its origin in the mother—a separate system that forms part of the kitten’s environment. This example is threatened by the simple expedient of re-bracketing the system under observation: if the system is the family of cats, the source of the action now lies within the bounds of the system, the identity and normativity conditions are met, and it seems that the act is now agentive in the strong sense the authors seek. I do not wish to argue that the authors are correct or incorrect in the attribution of agency. I wish to point out that the coherence of the notion depends upon the framing of the observation. To the extent that the family can be considered a system with its own dynamic identity, it is available as a possible locus of agency<sup>40</sup>. One of the core insights of the enactive approach is that autonomous organization is not fixed at one level. In mutual interaction, pluralities of autonomous systems can form novel superordinate emergent domains that themselves exhibit autonomy. With autonomy arises the need to consider the perspective of the emergent autonomous system, and the phenomenology enacted thereby. This general approach can be taken in considering interactions among cells, producing multicellular phenomena, including multicellular organisms. It can be applied to the emergence of social phenomena, e.g. riots or Mexican waves, when people engage in rich, reciprocal interaction with each other and within the confines of just the right set of constraints. And it can be applied to a mother cat and her kittens as a whole.

Related objections arise in the case of Parkinsonian tremors. Here, it is claimed, the human body is the system to be considered, the system is the active source of the interaction, but the normativity condition is not met as the tremor does not serve to maintain, protect, or establish any internally generated norm. Elsewhere, the argument is made that “the spasms of a person from Parkinson’s disease are not considered to be the [sic] actions, even though the person is a well-identifiable entity and the genuine source of her interactions with the environment” (Barandiaran et al. 2009: 5). The system has been variously identified here as the body, and the person. Neither seems entirely appropriate. Tremor arises when the equilibrium dynamics of a somatic subsystem changes from a static equilibrium to an oscillatory regime.

---

<sup>40</sup> Perhaps the mother + kittens is a poor candidate for the ascription of autonomy in the first place. In that case, my argument in this instance is weakened, but the general observations about the role of the framing of the system remains.

The fact that an oscillatory dynamic arises at all is evidence that there is indeed a system underlying the phenomenon, but it is not the whole body. The persistent copulatory motions of a male preying mantis after decapitation illustrates the notion of an oscillatory regime in a system that is less than a whole body rather well.<sup>41</sup> The assignation of the tremor to the body is thus misplaced, but the confusion of the body and the person is a far more serious matter, to which we will return in the next section.

The final example adduced by Barandiaran et al. (2002) is that of passive osmosis at the boundary of the cell. Here, the system seems to be uncontroversially the cell, and if the osmotic process subserves metabolism, the normativity condition is met. Agency, it is claimed, is not present because the system (the cell) is not the “active” source of the interaction. This example seems to typify a deep linguistic problem when discussing agency. Many verbs of action implicitly lean upon the background notion of an actor. Thus if we say that a property  $p$  of a system  $s$  is modulated, it is hard to make sense of that statement unless there is an implied agent, the modulator, who is doing the modulation. The asymmetry condition employed in the definition of agency in Barandiaran et al. (2009) seems to become circular for this reason when they say:

We therefore define interactional asymmetry as the condition describing a system as capable of engaging in some modulations of the coupling [between system and environment] and doing so at certain times. (Barandiaran et al. 2009: 4)

If a system “engages in some modulations”, we might well view it as an agent, but the presumption of agency lies here in the very description that employs an agentive verb. Such circularity infests the literature on agency. Kauffman (2002: 49) defines an agent as a system doing something on its own behalf. The verb “doing” carries the implication of a do-er, and circularity threatens again.

### 3. The Eye of the Beholder

Both asymmetry and normativity, it seems, depend upon the framing of the phenomena by an observer. In this, they have a lot in common with the individuality criterion itself. As Rohde and Stewart noted, taking the status of the observer seriously “transforms our conceptual world in a way that blurs the boundaries of what we normally consider a belief and what we consider a fact” (Rohde and Stewart 2008: 425). This groundlessness that necessarily arises when one eschews a positivist commitment presents something of

---

<sup>41</sup> If truth be told, this example comes from a Tom Waits song, and not a detailed study of the entomological literature.

a challenge if claim and counterclaim are to be judged on criteria other than opinion. The pragmatic approach suggested in Rohde and Stewart (2008) for treating of autonomy *simpliciter* and followed by Barandarian et al. (2009) is to shore up the justification of one or other perspective by the postulation of generative mechanisms. In so doing, they seek to go beyond mere description, while stopping short of an inflexible essentialism. But the ability to characterize a generative mechanism, by itself, does not achieve what the authors seek. It is possible to provide algorithmic effective procedures to illustrate many kinds of process, but that does nothing to establish the accuracy of the mapping from the algorithm to the world. This is a rhetorical strategy that has, in the past, lent the domain of artificial intelligence and some varieties of cognitive psychology a veneer of objectivity without due warrant. For example, proponents of a very different kind of agenda have sought to justify the creation of elaborate representational mental models using the same argument:

The theory should be describable in the form of an *effective procedure*...If a procedure can be carried out by a simple machine, plainly it does not require any decisions to be made on the basis of intuition or any other such 'magical' ingredient: it is an effective procedure. (Johnson-Laird 1983: 6)

By avoiding appeal to intuition, or even magic, the resulting theory is cast as objective, and hence inherently trustworthy. But the argument does not even try to ascertain the appropriateness, viridicality, or utility of the relation that obtains between the elements of the theory and the nature of the phenomenon being studied. So the appeal to generative mechanisms may serve to convince friends, but will hardly convert sceptics.

An alternative perspective on our strong predilection for attributing agency to one system but not another is provided by the claim that life can only be known by life. Because we are ourselves beings whose existence is a continual striving, we recognize this striving in others. If we were disembodied intellects, no such communion with the living would make sense, and the concepts of *organism* and *agency* would not have any grounding. Thompson argues:

To make the link from matter to life and mind, from physics to biology, one needs concepts like *organism* and *autopoiesis*, but such concepts are available only to an embodied mind with firsthand experience of its own living body. (Thompson 2004: 90)

This argument has been frequently made, and is sometimes attributed to Hans Jonas (1968). It is unclear to me whether both of the concepts of autonomy and agency should be considered to be members of the set of concepts that can only be understood by an embodied living being. It does not seem to stretch credulity too much to assume that tornados and flames might be recognised as examples of dynamically individuated phenomena to an abstract or alien intelligence with the ability to make observations at a human-like time and spatial scale. Jonas dismisses their individuality on the grounds of

intuition alone (Jonas 1968: 240), but Jonas is not drawing a distinction between the organisational closure and dynamical persistence that characterises autonomous organisation on the one hand, and the striving of an agent exhibiting behaviour regulated by goals or norms on the other.

There is thus a tension in the literature that informs the enactive perspective. On the one hand, Maturana (and possibly Varela) insists that:

Purpose or aims ...belong to the domain of our discourse about our actions, that is, they belong to the domain of descriptions, and when applied to a machine, or any system independent from us, they reflect our considering the machine or system in some encompassing context...Accordingly, if living systems are physical autopoietic machines, teleonomy becomes only an artifice of their description which does not reveal any feature of their organization, but which reveals the consistency in their operation within the domain of observation. Living systems, as physical autopoietic machines, are purposeless systems. (Maturana and Varela 1991: 85–86)

On the other, Jonas objects that biological organisms simply are not machines:

[W]hen we call a living body a “metabolising system,” we must include in the term that the system itself is wholly and continuously a result of its metabolising activity, and further that none of the “result” ceases to be an object of metabolism while it is also an agent of it. For this reason alone, it is inappropriate to liken the organism to a machine...food is more than fuel...its role is to build up originally and replace continually the very parts of the machine. Metabolism thus is the constant becoming of the machine itself—and this becoming itself is a performance of the machine: but for such performance there is no analogue in the world of machines. (Jonas 1966: 76)

Given the degree to which Varela leans upon the work of Jonas later (e.g. in Weber and Varela 2002), there does not seem to be a fundamental disagreement here, but rather an occasional failure to knowingly distinguish between the domain of description, as things appear to an observer who is, herself, an agent, and the domain of operation of the system. Recognizing this, it seems to me, opens up opportunities to further our understanding of autonomy and agency in new ways.

#### **4. Webs of Meaning**

As living entities, we are enmeshed in webs of significance and meaning. The tantalising prospect opened up by the enactive approach is to lay the foundation for a rational, scientifically informed account of our goings-on that can lean upon such notions as value and meaning, without descending into mere relativism and the brandishing of opinions. But this must be done self-consciously. When we speak of a value, it is with respect to some system for which we recognize a degree of autonomy. We can ground such discussion through observation of the degree to which a given system exhibits organisa-

tional closure, by examining the manner in which the system interacts with its immediate environment. We can do this to understand the conditions under which the autonomous identity exhibited by the system will persist, and the conditions under which it will be threatened.

The construct of the person and the organic reality of the human body need to be clearly distinguished here. Some behaviours exhibited by my movements make sense when referred to the domain of the body, as, e.g. when I recoil from a physical threat, seek water to still my thirst, or sleep. Very many of my behaviours do not make sense with respect to the somatic domain alone. When I go to a movie, play a tennis match, or vote in an election, these actions are structured by value, but the interpretation of the behaviour as subserving goals arising from one or other domain is not straightforward. There is no guarantee that an activity that can be picked out linguistically ("playing a game of tennis") can be interpreted with respect to the values arising from any single domain. If the game of tennis is a doubles match, we can recognize the domain of my body, the dyadic domain that is my team, and the set of four players who together enact the game of tennis. Some features of the activity are best understood with respect to each of these three domains, and there are probably further domains that one could sensibly identify. Furthermore, both the somatic individual, and the dyad that is a team, may admit of agentive interpretations, though I see no such interpretation for the set of four players.

Where an agentive ascription seems appropriate, the language of intentionality, of goal directedness and striving, can sensibly be applied. The desires, intentions and sensibilities of a group of protesters chanting in unison seem, to this author, to be no more or less real than the desires, intentions and sensibilities of me on the tennis court trying to win a match. Intentional predicates are easily and naturally used with respect to institutions, nation states, mobs, armies, and teams. On a conventional psychological reading, these uses are metaphorical extensions of the one true sense grounded in the (Cartesian, solipsistic) mind of the cherished individual. But from an enactive perspective, we can be flexible not only with respect to the identification of domains that exhibit temporally extended dynamic identity, but also with respect to the implied subject, that provides the values and normative scaffolding that allows mere movement to be interpreted as action or behaviour.

This constitutes something of a reversal of the normal course of scientific inquiry. The maxim introduced earlier encourages us to constantly ask both what do we see, and what is the implied subject of that seeing. This groundlessness is well known within the traditions of Buddhist and Taoist Philosophies, but has only rarely been acknowledged as a valid epistemological stance within the domain of science (Varela et al. 1991). In ascribing agency, we are tacitly acknowledging a commonality with the system being observed—

not a commonality based on the human body, but based on the lived experience of striving that is at the center of the experiencing subject.

### 5. Disinterest and Empathy

There is here an opportunity to develop a principled manner in which intentionality and value can be rationally treated within our collective discourse. Where the received psychological tradition insists on a single, immovable, split between subject and world, we are now in a position to recognize that any such split lies within the domain of description of this phenomenon or that. As our discourse ranges over the affairs of the body, of the family, of the schools and institutions of society, of the relation between man and climate systems, and beyond, the ground beneath the discussion shifts, and agency, attributed now this way and now that, brings into being first this subject and now that. Some of these subjects will appear singular, but some speak of collective intentionality, collective values, values originating with respect to an ever-changing topicalised object.

I bemoaned earlier the casual description of the domain underlying a Parkinsonian tremor, first as the body, and then as the person. Neither is the appropriate domain to ground the observation of the tremor, but the confusion of the body and the person is a far more serious confusion. We can speak with a degree of disinterest of the body. For many of us, we can contemplate replacing limbs, organs, and the like, with artificial prosthetics, without feeling that the person associated with the body is thereby fundamentally changed. For when we speak of the person, we speak of the value-laden ground of experience, of subject-hood. And the subject is neither singular nor plural. Fixing the subject at one level, as conventional psychological theory does, privileges some values over others, introduces a normativity that is utterly at odds with the richness and dynamic constitution of our being.

We can ascribe autonomy in a disinterested fashion, and we should do so. In this way we can distinguish between the chimera of the transient and the persistent identity of the autonomous system. When we go further and ascribe agency, we take sides, and express a degree of empathy, however slight. In the agentive description we acknowledge some relations between a system and its milieu as privileged, and accord them value. In dealing with the goings on of the animate, we have no choice but to traffic in the values of the animate. With life, value leaks in.

We must, therefore, be careful not to conflate the distinct notions of autonomy and agency. Autonomous organisation can be recognised in systems that do not display agency, as well as those that do. The ascription of autonomy in the sense of organisational closure or dynamic identity is done based on a set of criteria that are rooted in our observations. Changing the timescale, spatial

scale, or granularity of our observation may lead us to recognize that any ascription of autonomy is context-dependent and is done self-consciously by us as observers. The ascription of agency goes much further, and leads to the recognition of value—not an unprincipled recognition of value, but one grounded in system identity. However as observers, we become more firmly entwined with the objects of our observation when we engage in discussion of norms, goals, and values. These discussions must be done with an even greater degree of care.

There are differing views about the future utility of the enactive perspective. In some respects, it does not sit neatly within the contingent administrative divisions that characterize our institutions of learning and research. It is not psychology, nor is it merely social science, nor biology, and heaven forbid, it should not be mere philosophy. There are very many issues that are well addressed within the conventional frameworks, and for those there is no urgency in insisting upon an alternative approach. A cardiologist dealing with an acute heart attack does not need to ponder whether the function of the heart is better referred to the domain of the body or the domain of the description of the body, or just what the domain of the subject here is. The rehabilitation worker who needs to address the difficult re-integration of the post-heart-attack patient into the overlapping and competing spheres of family and occupational life, on the other hand, might have need of some principles to guide that complex task.

The advantage to be gained from adopting a fundamentally new perspective is that questions may now be posed, and perhaps addressed, that lie beyond our present competences. It is therefore to be hoped that the careful curation of the enactive vocabulary may help us to untie the conceptual knots that arise from a sterile and immovable split between mind and world. Here the ill-structured theoretical quagmire of mental health appears as an obvious domain of potential application. But so too do numerous societal issues that bear upon the relation of the individual to the many forms of collectivities that together make up our lives. Thinking further ahead, the collective that is humanity will inevitably face major issues for which it is ill-prepared, and that demand working solutions couched in a vocabulary of a collective that is neither imposed nor fictitious. The “we” will be negotiated, and that is only possible if we are capable of an ontological light touch in the matter of the (many) subject(s).

## 6. Acknowledgements

Marek McGann contributed mightily to lengthy and ongoing discussions around these themes, and while he cannot be held responsible for anything herein, if it has merit, some of that lies with his contributions. The little-known interpreter of Buddhist and Taoist texts who wrote under the name of Wei Wu Wei also lurks unacknowledged in the undergrowth of this article.

## References

- Barandiaran, X. E., Di Paolo, E., and Rohde, M. 2009. Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, 17(5): 367–386.
- Beer, R. D. 1995. A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72(1): 173–215.
- Christensen, W. and Hooker, C. 2000. Autonomy and the emergence of intelligence: Organised interactive construction. *Communication and Cognition-Artificial Intelligence*, 17(3-4): 133–157.
- Deacon, T. W. 2011. *Incomplete Nature: How Mind Emerged from Matter*. WW Norton & Company.
- Dennett, D. C. 1989. *The International Stance*. The MIT press.
- Johnson-Laird, P. N. 1983. *Mental models: Towards a Cognitive Science of Language, Inference, and Consciousness*, volume 6. Harvard University Press.
- Jonas, H. 1966. *The Phenomenon of Life: Toward a Philosophical Biology*. Northwestern University Press.
- Jonas, H. 1968. Biological foundations of individuality. *International Philosophical Quarterly*, 8(2): 231–251.
- Kauffman, S. A. 2002. *Investigations*. Oxford University Press.
- Maes, P. 1993. Modeling adaptive autonomous agents. *Artificial Life*, 1(12): 135–162.
- Maturana, H. R. and Varela, F. J. 1991. *Autopoiesis and Cognition: The Realization of the Living*. Boston Studies in the Philosophy of Science. Springer.
- Moreno, A. and Etxeberria, A. 2005. Agency in natural and artificial systems. *Artificial Life*, 11(1-2): 161–175.
- Murphy, J. P. and Rorty, R. 1990. *Pragmatism: From Peirce to Davidson*. Westview Press Boulder, CO.

- Rohde, M. and Stewart, J. 2008. Ascriptional and 'genuine' autonomy. *BioSystems*, 91(2): 424–433.
- Ryle, G. 1949. *The Concept of Mind*. University of Chicago Press.
- Stewart, J., Gapenne, O., and Di Paolo, E. A. 2010. *Enaction: Toward a new Paradigm for Cognitive Science*. The MIT Press.
- Thompson, E. 2004. Life and mind: From autopoiesis to neurophenomenology. a tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences*, 3(4): 381–398.
- Varela, F., Thompson, E., and Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. MIT press.
- Varela, F. J. 1979. *Principles of Biological Autonomy*. North Holland New York.
- Weber, A. and Varela, F. J. 2002. Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1(2): 97–125.