

Investigations of an Adaptive and Autonomous Sensorimotor Individual

Matthew Egbert

University of Auckland, Auckland, NZ
mde@matthewegbert.com

Abstract

Agency is an under-investigated foundational concept in understanding natural minds and how they differ from existing artificial forms of intelligence. To address this, Barandiaran et al. (2009) outlined a provisional definition of minimal agency, based upon three criteria: autonomous individuality; asymmetrical agent-environment interaction; and norm-driven modulation of that interaction. The first part of this paper reviews this definition, drawing attention to the interaction between interactional asymmetry and normativity. The definition is then applied to self-maintaining sensorimotor dynamics observed in a computational model. This has two broad goals: (i) improving our understanding of Barandiaran et al.'s definition of agency and how it could be applied to sensorimotor dynamics; and (ii) improving our understanding of the agent-like structures observed in a simulation of a simple robot whose sensors and motors are coupled to an iterant deformable sensorimotor medium (IDSM). I argue that specific structures within the simulation qualify as autonomous individuals and that these individuals can adapt to environmental changes in a way that benefits their viability. The nature of this adaptation is then examined by comparison to metabolism-independent and metabolism-based form of bacterial chemotaxis.

Introduction

Since its inception, mainstream artificial intelligence research has focused upon knowledge, representation and problem solving. The dominating cognitivist and connectionist approaches have produced a number of remarkable technological advances, to which they continue to contribute. However, a number of central mental phenomena, such as subjective experience, intrinsic goals and qualia, have proven largely inaccessible to these frameworks, leading to a number of rather extreme responses, including the denial or dismissal of our subjective experience as an illusion (Dennett, 1992); a resurgence in mind-body dualism, where mental phenomena are seen as outside the realm of scientific explanation (Chalmers, 1997); and panpsychist theories that suggest, for instance, that everything is conscious, but varies in the extent to which it is conscious depending upon the amount of 'integrated information' (Tononi, 2004).

By comparison, the enactive view (Stewart et al., 2010; Thompson, 2007; Varela et al., 1992) is rather conservative. Enactivists reject abstract and disembodied problem-solving as the theoretical foundation for understanding mind. Instead, mind is investigated as a property of a particular 'autonomous' organization of interdependent, precarious components. I elaborate upon this idea below, but the key point here is that where mainstream AI researchers arbitrarily define the agents that they are interested in—a neural network, piece of software or robot that solves a particular problem—enactivists have proposed stricter criteria for what is necessary to consider a system an agent. These criteria pre-empt the arbitrary assignment of agency to a system of interest, but with this constraint comes a potential benefit in the form of a line of investigation for improving our understanding of the mental phenomena that have been so elusive to mainstream cognitive sciences and AI.

In this vein, Barandiaran, Di Paolo and Rohde (2009) provide a preliminary definition of agency. The goal is that once fully developed, this definition could be applied to identify within a given system, the presence of agents and the boundaries thereof. In this paper, I apply this definition (henceforth BDR-Agency) to a model of a sensorimotor agent. Our goal is to improve our understanding of both the self-maintaining sensorimotor dynamics observed in our model and of the definition itself. BDR-Agency is based upon three properties: individuality, normativity, and interactional asymmetry. I lack space to recapitulate these in full detail, but shall provide a brief overview before introducing our model of a sensorimotor agent.

The foundational concept is *autonomy-based individuality*, where an individual is seen as a set of interdependent processes or components. To be a member of the set, each constituent process must depend upon at least one other constituent process, and must make possible at least one other constituent process (Ruiz-Mirazo and Moreno, 2004; Varela, 1979; Virgo et al., 2009). In this view, it is not the location of a component that determines whether it is part of an individual, it is how that component relates to (i. e., does or does not enable and/or depend upon) the rest of the individual.

This often maps fairly well to our intuitive understanding of a biological organism, but not always. For example Di Paolo (2009) describes how in this view, the pockets of air that are trapped in the hairs of the water boatman insect (*Corixidae*) are *part* of the organism as they are essential for and are maintained by the other processes that constitute the creature.

Autonomy provides a principle-driven way to identify individuals (and consequently, their environment). It, like the other criteria in BDR-Agency, is a domain-independent organizational property, meaning that its application is not limited to identifying and delimiting biological individuals (organisms); but can also be applied to identify individuals in other domains including social, ecological, and sensorimotor systems. What constitutes an individual is no longer an arbitrary decision, opening up new meaningful lines of enquiry concerning the properties of such individuals and the relationship between individuals and their environment. In this vein, BDR-Agency proposes additional criteria for an autonomous individual to be considered an agent: it must modulate its interaction with its environment in response to the norms that are intrinsic to its organization. Let us now briefly unpack these ideas.

Given an autonomous individual, it is possible to evaluate phenomena with regard to how they affect the individual's viability. This is a *normative* (good/bad) evaluation that is determined by (i. e., is the result of) the organization of the autonomous individual. To explain: if the individual were to be built out of different components, or the same components were organized in a different way, the set of phenomena that influence its survival, and the way that they influence its survival would be different. This normative evaluation is thus *not* something that is decided by an external observer of the system, but rather is determined by the individual's organization. It is as such intrinsic rather than ascriptional.¹

Having outlined what is meant by intrinsic norms, I can turn our attention to the notion of an autonomous individual modulating its interaction with its environment in response to these norms. A fundamental concept in agency is the difference between the actions of an agent and the things that just happen to it. The intuitive idea is that when an agent acts, it has played a greater role in causing something to happen than the environment, but the challenge of quantifying this asymmetric causal contribution has a long history and has proven difficult. After briefly reviewing some energetic, dynamical and information-theoretical approaches to resolving this topic and some problems as of yet unresolved in these approaches, Barandiaran et al. propose that the autonomy based concept of individuality provides a way "to define interactional asymmetry in terms that are weaker than

those of causation, but also less problematic." (Barandiaran et al., 2009, p. 371) To do so, they draw attention to the interaction between the agent and its environment, suggesting that the agent influences this interaction in a way that is different from how the environment influences it. They start by rightly observing a symmetrical relationship of mutual influence between individual and environment, which they express by writing $dS/dt = F_Q(S, E)$; $dE/dt = G_Q(S, E)$, where S is the state of the individual and E the state of the environment. Then, they suggest that the form of these functions (F_Q and G_Q) is determined by a set of parameters Q , and highlight p , the subset of Q that is influenced only by S : $\Delta p = H_T(S)$, (where the subscript T is used to emphasize that the modulation is an event, not a lawful, general rule of change), using this to express the idea that "an agent is able to modulate some of the parametrical conditions and to constrain this coupling in a way that the environment (typically) does not." (Barandiaran et al., 2009, p. 371)

After emphasizing that individuality is a precondition for the other two criteria, Barandiaran et al. (2009, pps. 373–374) present interactional asymmetry and normativity as independent properties, giving examples of systems that provide one but not the other. Here I want to focus upon the relationship between these properties. Instead of considering them as independent, let us consider the agent-environment interaction as asymmetrical *because it is done in response to the norms of the agent?*² When a mechanism is observed through which an autonomous individual influences its interaction with its environment, the question can be asked what is modulating this influence—what causes the form of this influence to change? If and only if it is in an adaptive (*sensu* Di Paolo, 2005) response to one of the processes that constitute the individual, then it is the action of an agent (cf. a mere happening).

Parkinson's tremors are not varied in an adaptive, survival-benefiting response to one or more of the processes that constitute the agent, and so they are not the acts of an agent. The metabolism-based chemotaxis of various bacteria (Alexandre, 2010; Egbert et al., 2010), on the other hand, is agential as it is directly modulated by the efficacy of the (metabolic) processes that are necessary for the individual to persist. Not all bacterial chemotaxis operates in this manner. Metabolism-independent chemotaxis (which is also observed in a variety of bacteria—see e. g. Adler, 1969) is modulated in response to particular environmental changes—i. e., not in response to the efficacy of the processes that constitute the agent. When, for instance, a mutation blocks the consumption of that resource, the metabolism-independent behaviours do not change or cease; and the organism continues to move toward the 're-

¹This approach to naturalizing norms follows the Kantian proposal that norms are to be derived from what makes possible the existence of certain systems, rather than a more problematic notion of what *ought* to be—see (Barandiaran and Egbert, 2013, p. 7).

²Di Paolo et al. (2017) present an updated, more integrated view similar to that described here where the three requirements are not independent or additive, but relate non-trivially to each other as argued here.

sources’ that it no longer needs, even when this means its death (Adler, 1969). A contentious proposal would be to suggest that despite the tendency of this mechanism to benefit the organism, because it is not modulated in response to processes that constitute the organism it is *not* an action! Di Paolo et al. (2017, pps. 131–132) do not go this far, but do consider ‘lucky-accidents’ (incidental benefits that result as a side-effect of mechanisms that are not attuned to those benefits) to not be actions. Metabolism-independent chemotaxis is different, they argue, because it is the result of evolution that is constrained by the norms of the organism (Di Paolo, personal communication).

After outlining their definition, Barandiaran et al. show how it can be applied to the archetype which played a role in inspiring their ideas: bacterial chemotaxis. Di Paolo et al. (2017) consider the concept in another domain, investigating how certain sensorimotor dynamics could satisfy the requirements of BDR-Agency. These are useful efforts to consider how BDR-Agency applies to natural systems, but it can also be useful to develop and analyse computational artefacts that attempt to capture key properties in minimal ways. In collaboration with some of the authors listed above, I have used this methodology to investigate and help refine concepts of normativity (Barandiaran and Egbert, 2013), and norm-driven behaviour (Egbert et al., 2010, 2009). In what follows, I present a computational model that we use to further consider how BDR-Agency can be applied to sensorimotor dynamics. The model is much simpler than biological systems, and more formal and concrete than purely verbal conjecture, and thus provides a supplementary philosophical method, that can shed additional light upon BDR-Agency.

Model

The Iterant Deformable Sensorimotor Medium (IDSM) is a dynamical construct, that when coupled to a robotic or simulated body, maps the current sensorimotor state to a motor output $f_t(S_t, M_t) \rightarrow M_{t+1}$. As it is applied, the mapping itself changes as a function of the current state of sensors and motors and the current state of the mapping: $\frac{\Delta f}{\Delta t} = g(f, S, M, \frac{\Delta S}{\Delta t}, \frac{\Delta M}{\Delta t})$. This change (detailed below) was designed so that sensorimotor state trajectories bias the system to increase the likelihood that similar sensorimotor trajectories will be repeated in the future. This property is reminiscent of the self-reinforcing nature of habitual behaviour, where repeated performance of patterns of behaviour (e. g. the direction you look when crossing the street, or a tendency to worry) increases the likelihood of similar behaviour being performed in the future. A useful metaphor for understanding how the IDSM works is the stigmergic path generation observed on university campuses, where routes taken by students crossing a field between academic buildings trample paths in the grass that influence the trajectories taken by subsequent students. This is essentially how the IDSM operates, but the trajectories and stigmergic

paths are in sensorimotor space, rather than on grass.

Now for more formal detail: the IDSM tracks a history of sensorimotor trajectories by creating and modifying a set of records known as *nodes*. Each node describes the sensorimotor-velocity (i. e., the rate of change in all sensors and motors) for a particular sensorimotor-state at some point in the past. Each node is thus a tuple of two vectors and a scalar, $N = \langle \mathbf{p}, \mathbf{v}, w \rangle$, where \mathbf{p} represents the sensorimotor state associated with the node (referred to as the node’s position in sensorimotor space), \mathbf{v} indicates the sensorimotor velocity, and the scalar w indicates the weight of the node, a value that scales the overall influence of the node. We shall refer to these components using a subscript notation, where the position, SM-velocity vector, and weight of node N are written as $N_{\mathbf{p}}$ and $N_{\mathbf{v}}$ and N_w , respectively.

As a robot controlled by the IDSM moves through sensorimotor states, new nodes are created recording the sensorimotor velocities experienced at different sensorimotor states. Specifically, when a new node is created, its $N_{\mathbf{p}}$ is set to the current sensorimotor state; its $N_{\mathbf{v}}$ is set to the current rate of change in each sensorimotor dimension, and its N_w is set to 0 (an initial value that does not imply that the node is ineffectual, see below). The two vector terms ($N_{\mathbf{p}}$ and $N_{\mathbf{v}}$) are calculated in a normalized sensorimotor space, where the range of all sensor and motor values are linearly scaled to lie, in each dimension in $[0, 1]$. In similar models in the past, the sensorimotor state could become trapped at the corners of the sensorimotor space (Egbert and Barandiaran, 2014). To compensate, I made the motor dimension of sensorimotor space periodic and mapped the motor sensorimotor coordinates by a sinusoidal function so that $N_{\mathbf{p}}$ and $N_{\mathbf{v}}$ coordinates are mapped to motor values according to $m = \sin(2\pi \cdot x)$, where x is the motor coordinate in SM-space and m is the actual motor value.

New nodes are added only when the weighted density of nodes near the current sensorimotor state, as described by the function ϕ (Eqs. 1–3), is less than a threshold value, $\phi(\mathbf{x}) < k_t = 1$. Loosely speaking, ϕ is a measure of how familiar the sensorimotor state is, and it is calculated by summing a non-linear function of the distance from every node to the current sensorimotor state, $d(N_{\mathbf{p}}, \mathbf{x})$, scaled by a sigmoidal function of the node’s weight, $\omega(N_w)$, thus:

$$\phi(\mathbf{x}) = \sum_N \omega(N_w) \cdot d(N_{\mathbf{p}}, \mathbf{x}) \quad (1)$$

$$\omega(N_w) = \frac{2}{1 + \exp(-k_\omega N_w)}; k_\omega = 0.0025 \quad (2)$$

$$d(N_{\mathbf{p}}, x) = \frac{2}{1 + \exp(k_d \|N_{\mathbf{p}} - \mathbf{x}\|^2)}; k_d = 200 \quad (3)$$

After a node is created, its weight changes according to equation 4, where the first term represents a steady degradation of the node’s influence, and the second term represents a

strengthening of the node that occurs when the current sensorimotor state is close to the node’s position. This latter term allows for the self-reinforcement/self-maintenance of patterns of behavior, such that patterns of behavior that are repeated are more likely to persist than those that are not reinforced.

$$\frac{dN_w}{dt} = -k_{deg} + r(N, \mathbf{x}); \quad k_{deg} = 5 \quad (4)$$

$$r(N, \mathbf{x}) = k_{rejuv}d(N_{\mathbf{p}}, \mathbf{x}); \quad k_{rejuv} = 1000 \quad (5)$$

One time unit after creation, nodes are activated, meaning that they are added to the pool of nodes that influence the motor state. Every activated node influences the motor state, but at any one time only a subset of these will have a substantial influence, for the influence of a node is scaled non-linearly by its distance from the current sensorimotor state by the same distance function used in ϕ above (Equation 3). The influence of each node is also scaled by its weight according to Equation 2, and thus nodes that are close to the current sensorimotor state and nodes with higher weights have a greater influence.

Equations 6 – 8 describe how the IDSM influences the motor state (μ). The influence of a node can be broken down into a ‘velocity’ factor and an ‘attraction’ factor. The velocity factor is simply the motor components of the $N_{\mathbf{v}}$ vector, but the attraction factor is slightly more complicated. It is a ‘sensorimotor-force’ that draws the system towards the node. This tends to result in a motion in sensorimotor space towards regions of sensorimotor space that are familiar, i.e. for which there is a higher density of nodes and it can compensate for stochasticity in the environment or perturbations to behavior (see Egbert and Barandiaran, 2014 for details).

$$\mathbf{I} = \frac{1}{\phi(\mathbf{x})} \sum_N \omega(N_w) \cdot d(N_{\mathbf{p}}, \mathbf{x}) \cdot \underbrace{\left(\frac{N_{\mathbf{v}}}{\text{Vel.}} + \underbrace{A(N_{\mathbf{p}} - \mathbf{x}, N_{\mathbf{v}})}_{\text{Attraction}} \right)^\mu \quad (6)$$

$$A(\mathbf{a}, N_{\mathbf{v}}) = \mathbf{a} - \left(\mathbf{a} \cdot \frac{N_{\mathbf{v}}}{\|N_{\mathbf{v}}\|} \right) \frac{N_{\mathbf{v}}}{\|N_{\mathbf{v}}\|} \quad (7)$$

In previous IDSM publications, random behaviour was generated by creating a large number of initial random nodes. In this paper I have replaced this by having the local density of nodes, $\psi = \sum_N d(N_{\mathbf{p}}, \mathbf{x})$, determine when IDSM influence controls the motors and when the motor activity is to be random. Eq. 8 expresses how a sigmoidal switch ($s(\psi)$) selects either IDSM driven behaviour (I) or a random motor vector (R). R is varied over time to produce a random walk in motor space as follows: 100 times per time unit, R has a 0.02 chance of being set to a vector with motor-components selected from a normal distribution ($\sigma^2 = 2$), and a delta component, ρ , selected from a normal distribution ($\sigma^2 = 10$), such that $dR/dt = \rho$.

$$s(\psi) = \frac{1}{1 + \exp(20\psi - 20)}; \quad \frac{d\mu}{dt} = (1 - s)\mathbf{I} + s\mathbf{R} \quad (8)$$

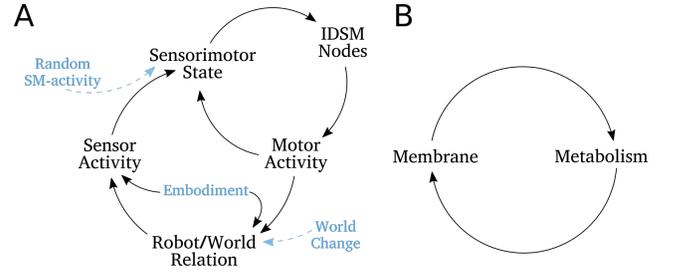


Figure 1: Dependencies between the precarious dynamics that constitute (A) sensorimotor and (B) biological identities. Black arrows indicate ‘is necessary to maintain’ and the black components they interconnect constitute a sensorimotor individual. Dashed arrows indicate a weaker ‘influences’ relationship, and the blue text indicates components that are not part of the individual as they do not depend upon the other components, and are thus instead part of its environment. See the main text for details.

As discussed in previous publications, when coupled to a robot’s sensors and motors, the IDSM (i) causes the robot to repeat behaviors that it has performed in the past, and (ii) allows for the reinforcement of patterns of behavior through repetition. If a pattern of behavior is not performed for a period of time, it becomes less likely to be re-enacted, but when behaviors are performed, they become more likely to be repeated in the future, and in this way, self-maintaining patterns of behavior emerge. It is these self-maintaining patterns of behaviour that I will investigate as sensorimotor agents.

Investigation

We now consider how BDR-Agency could be applied to structures within our model. We start by identifying an autonomous individual—i. e., a network of interdependent precarious entities. The model was designed to include such ‘sensorimotor individuals’ (henceforth SMI³), and so this is not a difficult task, but it nevertheless proves interesting. We can start by considering the influence of IDSM nodes, which perpetually decreases, except when the robot’s sensorimotor state is rather close to the node, in which case the weight of the node is rejuvenated (Eqs. 4–5). A given node will therefore cease to have influence unless the region of sensorimotor space near it is visited regularly. The influence of IDSM nodes is thus precarious in a way that is comparable to the precariousness of far-from-equilibrium metabolic process in a biological individual. They are inherently unstable; in the short term, their own influence drives them away from the conditions necessary for their survival (IDSM nodes tend to cause change in sensorimotor state, which will move it away from the node), but when certain conditions are met, these

³In previous publications, these autonomous structures have been referred to as ‘habits’, due the similarity in the way these structures depend upon their self-reinforcement to persist.

structures can persist indefinitely.

So, while precarious, IDSM nodes are not autonomous, as their influence in isolation tends to move the sensorimotor state away, causing it to degrade more rapidly. They can however be part of an autonomous individual when they are integrated into a wider network of precarious interdependent components as drawn in Fig. 1. Let us now trace these dependencies. IDSM nodes influence the motor state of the robot (Eqs. 6–8). Motor activity in turn influences the motion of the robot and thus how the robot relates to the world in which it is situated (e. g., how close the robot is to the light). The robot’s relation to its world determines its sensory state, and this, along with the motor state (and at times the random behaviour generation) determine the robot’s sensorimotor state. The sensorimotor state of the robot must regularly visit IDSM nodes if their influence is to persist, and so we are back to where we started. We have circumnavigated a network of interdependent components depicted as black text and arrows in Fig. 1 and in so doing, have identified an autonomous individual within the model.

Note that these components are extended in time. It is not the *instantaneous* motor activity that contributes to the robot/world relation, but the historical trajectory of motor activity. Similarly, it is not the instantaneous sensory state that is important, but its trajectory of time. It is generally not a single IDSM node that is part of the SMI, but a collection of nodes that tends to cause a pattern of motor activity and it is a temporally extended pattern of robot/world relations that produces the repeated sensory states that allow for the reinforcement of certain IDSM nodes, etc.

We can identify elements of this SMI’s environment by further considering the robot/world relation. Note that the motor activity influences, but does not completely determine the robot/world relation. Different trajectories of motor activity could cause the robot to move toward or away from a light in an infinite variety of ways, and other trajectories would cause the robot to maintain a particular distance, or perform potentially complex oscillations in front of the light, etc. But motion of the light (including it being stationary) plays an equal role, co-determining change in the robot/environment relation. Changes in the world of the agent, such as this light motion, influence part of the SMI but do not depend upon it. As such light motion is not part of the SMI but is instead part of its environment. Care is needed here because I am using the word ‘environment’ in a potentially confusing manner. A more everyday use of the word ‘environment’ would refer to the world in which the robot operates. But, here, because we are taking the approach proposed by Barandiaran et al. (2009), the individual is not an arbitrarily selected robot, but instead the system which satisfies the criteria outlined in BDR-Agency. In this case, the individual is not coextensive with the robot, but is rather the SMI, the network of interdependent sensorimotor dynamics just described, and its environment is everything

that it interacts with that is not part of it. I use ‘world’ to refer to the non-robot parts of the simulation (e. g. lights) and ‘environment’ to refer to the non-SMI parts of the simulation.⁴

Perhaps counter-intuitively, the robot’s embodiment in our model is also part of the SMI’s environment—Di Paolo et al. (2017) reach similar conclusions. Its persistence is assumed and is thus not dependent upon the robot’s behaviour (or indeed anything else within the model). It influences the form of the equations that specify how the motors affect the robot/world relation and how the robot/world relation determines the sensory activity, but it does not depend upon any other part of the SMI, and so is not part of it. For biological individuals (organisms), the situation is different. Biological self-construction (a.k.a. autopoiesis—presented in a gross simplification in Fig. 1B) depends upon the behaviour of the organism, to get food etc., and the behaviour depends upon the autopoiesis to create the structures that enable the behaviour (flagellum etc.). And so for biological organisms, the networks depicted Fig. 1A and Fig. 1B are mutually interdependent and we have an example of nested autonomous structures; a larger autonomous individual that when analyzed at a different scale, could be identified as two sub-individuals—a sensorimotor individual and a biological individual. I do not consider this further in this paper, but we have used IDSM models to investigate the relationship between biological and sensorimotor autonomous structures (Egbert and Cañamero, 2014), and plan to do more research in this direction.

The norms of a sensorimotor individual

Having identified a sensorimotor individual, we can consider the norms that emerge from its organization. We can ask: *What is good for the survival of an SMI and what is bad for it?* To answer this question, we now look at an instance of an SMI in a simulation. To do so, we couple an IDSM to a simulated robot, situated in a one-dimensional periodic environment of size 2. The robot has one motor which determines its velocity $m \in [-1, 1] = 2 \frac{dx}{dt}$, where the periodic variable $x \in [-1, 1]$ is the robot’s position. It also has one sensor which is activated according to $\frac{1}{(1+5d)^2}$, where d is the minimum distance between a light source and the robot in accordance with the minimum image convention for periodic boundary conditions. Figure 2 presents a time-series indicating the position of the robot and the light during our experimental run. In this simulation, the light source moves according to a predefined sequence, remaining stationary, moving ‘left’ at a velocity of $-\frac{1}{5}$, oscillating, moving ‘right’ at a velocity of $-\frac{1}{5}$, before repeating this sequence. After an initial short transient during the initial phase where the light

⁴An additional source of potential terminological confusion is the notion of sensorimotor-environment proposed by Buhrmann et al. (2013) which is a third and distinct idea from the environment discussed here.

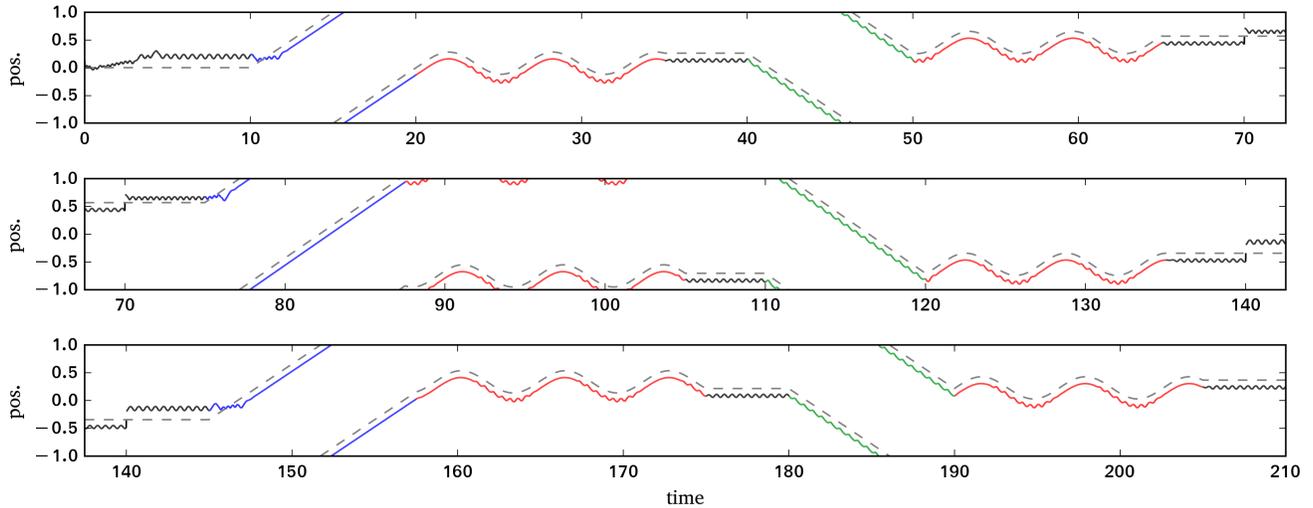


Figure 2: Time-series showing the motion of a light (dashed line) and simulated robot in a 1D periodic environment. Note the robot is teleported at $t = 70$ and $t = 140$ to lie on the opposite side of the light.

is stationary, the robot falls into a behaviour involving a low amplitude oscillation close to the light. When the light starts to move, the robot goes through another transient, changing from being ‘above’ the light to being ‘below’ it, where it remains maintaining approximately the same distance from the light for the remainder of the simulation.

The IDSM nodes that produce this behaviour are depicted in the lower-right plot of Fig. 3. Note that they are all located within a portion of sensorimotor space that has high sensory excitation (vertical axis). This makes sense as during this behaviour, the robot is always relatively close to the light. If we were to move the light away from the robot, the robot/environment relation would be changed, and this change, if sufficiently large, would cause the sensorimotor state to leave the area where (i) motor activity is influenced by the nodes and (ii) the nodes are reinforced. It follows that moving the light away from the robot in this way is bad for the survival of the SMI. The contrapositive is also true: if, after moving the light away from the robot, we move it back (in time that the IDSM nodes have not degraded too extensively), this second action would be *good* for the SMI. These examples (which have been observed in simulation but are not presented here) support the claim that different robot/world relations are better or worse for the persistence of the habit, and so change in robot/world relations takes on a normative dimension for the SMI. Other types of change can be similarly evaluated in terms of their effect upon the SMI’s viability. For instance it is evident that moving IDSM nodes or changing their various properties would often break a SMI. Similarly, modifying the robot’s embodiment would often break or at change the form of the trajectories that constitute the SMI.

Norm-sensitive modulation of agent-environment interaction by SMI

It is intuitive that by changing the system’s dynamics we can ‘kill’ an SMI. What is less clear is how robust or adaptable they are. Recall that we are interested in the extent to which SMI can regulate their interaction with their environment in response to their own emergent norms. A first step toward investigating this possibility is to evaluate the stability of the SMI in our model. There are a number of possibilities. It could be an unstable attractor—in this case, a perturbation no matter how small, would eventually destroy the SMI. Alternatively, the SMI could exist in a region of many proximal attractors where small changes permanently modify the SMI (or equivalently, destroy it and replace it with a different but similar SMI). Finally, the SMI could be a stable attractor with a basin of attraction such that the effects of small perturbations can be corrected for so that after some recovery time, the original SMI is restored.

The dynamics observed in our simulation suggest that this final possibility is the case for our SMI, at least when perturbed by variations in the light’s motion. In the absence of robot motion, changes in the motion of the light would produce changes in the robot/world relation. But, we have already observed that after the initial transient dynamics, the SMI maintains a consistent relationship with the light, despite the variation in the light’s motion. This suggests a certain degree of robustness to perturbation in the SMI, and recalling that the motion of the light is part of the environment of the SMI, we can argue that the SMI can adapt to different environmental conditions, maintaining the trajectory of agent-environment relations that allow the SMI as a whole to persist.

How does this adaptation occur? If we consider the IDSM nodes in the lower-right plot of Fig. 3, we can see that when

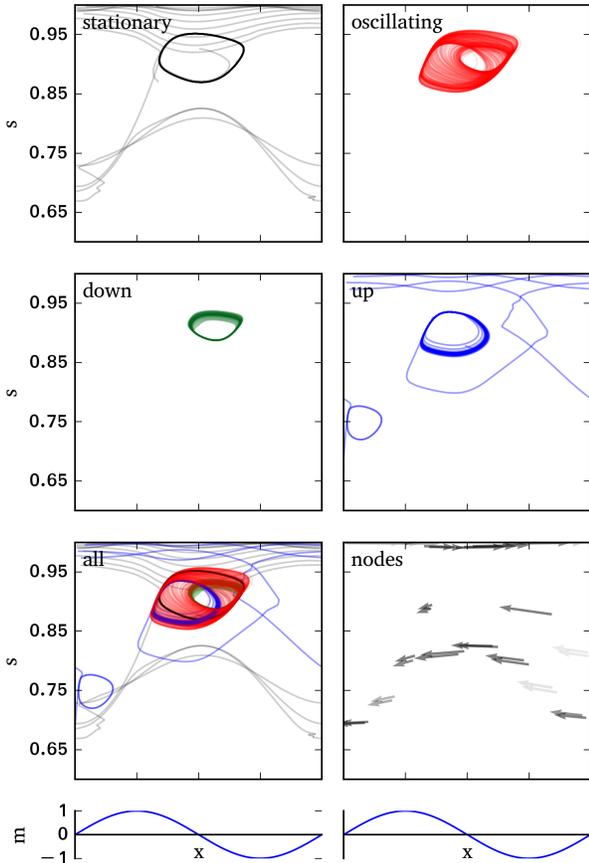


Figure 3: Sensorimotor dynamics of the robot during different motions of the light. The bottom-right figure shows the IDSM nodes involved in these dynamics. All dynamics shown are from before the experimental teleportation of the robot. The sinusoidal plots at the bottom indicate how the motor dimension of the sensorimotor space (x , the horizontal axis) maps to motor values (m).

the sensor value decreases, the motor-space coordinate (x) is decreased by the IDSM. The motor-coordinate \rightarrow motor value map (depicted in the bottom plots of Fig. 3) mean that this change causes an increase in m , which moves the robot closer to the light. The farther the sensor value decreases, the stronger the response (provided the sensor state does not go too far—say much lower than ≈ 0.8). The SMI is an emergent structure that connects sensors and motors in a negative feedback relationship that stabilizes that very structure. The regulation of the agent/world relation is thus normative, in that it is beneficial to the persistence of the SMI, but it is possible to make a stronger claim that the regulation is *a response to the norms of the SMI*, the way that we described above that metabolism-based chemotaxis is a response to self-constructing processes that constitute the chemotactic bacterium? In other words, is the ‘norm-following’ nature of this response coincidental, or is it a response to the ‘norm-generating’ processes—the efficacy of the processes that constitute the individual?

The sensorimotor state is an integral essential variable for the SMI. If it goes out of bounds (the ‘viability limits’), the SMI will likely die, but if it stays within this (and other) viability limits it will persist. As described in the previous paragraph, when this essential variable goes too low or too high, the SMI compensates. Paraphrasing Di Paolo’s definition of adaptivity (Di Paolo, 2005, p. 438), we have a case where trajectories toward the viability limits are transformed in such a way to prevent the viability boundary from being reached. As described above, the closer that this essential variable gets to the viability limits (e. g., the farther the sensory state is brought down) the stronger the response, and in this sense, this system seems to be responding to its own emergent norms.

However, when the agent is teleported to the opposite side of the light at $t = 70$, inverting the motor velocity needed to move toward the light, the SMI ceases to be able to compensate for changes, and the SM-state quickly leaves the SM-space associated with the SMI. The viability-limits have not changed, but the effects of the SMI that previously accomplished adaptation do so no longer. Instead of counteracting motion toward the viability limits, the SMI now accelerates such motion across the boundary, suggesting that the adaptation is not directly sensitive to its viability limits.

Nevertheless, the sensor activity is an essential variable of the SMI. It must remain within certain limits for it to survive and we have just described how the SMI compensates for certain trajectories that would otherwise cause the system to leave its viability limits. It seems reasonable to describe the adaptation as a rather direct response to the dynamics of one of the system’s essential variables, and accordingly it seems reasonable to propose that the adaptation is, in this sense, a rather direct response to the needs of the system. It is less clear however how one might consider this response as driven by the efficacy of its constituent processes. Further work is needed to explore these ideas.

One final comment about the robustness and longer-term adaptability of SMI is in order. After teleporting the robot, a new SMI was rapidly formed. The new SMI is an oscillation that looks similar in Fig. 2, but involves a combination of new nodes and some from the original SMI. When we subjected this new SMI to a similar set of environmental perturbations as before ($t \in [70, 140]$), it proved unstable, and rapidly ($t \approx 77$) the system returned to the initial SMI. When subjected to a similar perturbation ($t \approx 147$) the trajectories again returned to the original SMI, and the system as a whole has learned to accommodate the teleportation perturbation.

Discussion

In this paper, we have provided a computational model of an autonomous sensorimotor individual and provided a preliminary investigation of the limits of that individual (its constituent processes), its viability limits, and the norms that

emerge from its organization. We also explored the stability and robustness of SMI and the mechanism underlying its robustness. The model allowed us to explore how autonomous sensorimotor individuals could accomplish norm-following adaptation that is in direct response to norm-generating processes, but further work is needed in this area.

To overcome outdated views of mind, positive alternative accounts must be put forward. We live in exciting times as a number of such accounts are emerging and being given serious consideration. We should not sell these ideas short. It can be illuminating to see how other approaches in cognitive science such as predictive coding can be recast in enaction-friendly terms (Hutto and Myin, 2017), but we must go farther. One of the best ways to do so is to create artificial systems that instantiate the ideas.

The artificial intelligence community is committed to producing ‘problem-solving’ tools and as such it is obligated to the approaches that are *currently* the most useful at solving problems. This work benefits society, but if we are truly moving away from computationalist cognitivist functionalism, we may need to abandon (at least for a time) problem solving as a focus. The artificial life community is known for building abstract models that are used to investigate ideas and definitions rather than to maximise utility or simulate specific natural systems. ALife modelling can play a very useful role in helping to communicate, clarify, and simplify the difficult concepts and we hope that enactivists and others that are developing new ideas in cognitive science will increasingly engage with this kind of research.

In this paper, the IDSM-based model provided an example system that we could use to test and evaluate BDR-Agency. In our investigation we found individuals constituted by both ‘internal’ (to the robot) dynamics and robot-world relation dynamics, providing an example of autonomous individuals that are (as argued in Di Paolo, 2009) compatible with notions of the extended mind (Clark and Chalmers, 1998). In the model, we identified adaptive sensorimotor individuals, whose behaviour was not only normative, but arguably norm-driven.

References

- Adler, J. (1969). Chemoreceptors in Bacteria. *Science*, 166(3913):1588–1597.
- Alexandre, G. (2010). Coupling metabolism and chemotaxis-dependent behaviours by energy taxis receptors. *Microbiology*, 156(8):2283–2293.
- 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.
- Barandiaran, X. E. and Egbert, M. D. (2013). Norm-Establishing and Norm-Following in Autonomous Agency. *Artificial Life*, 20(1):5–28.
- Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72(1-2):173–215.
- Beer, R. D. (2015). Characterizing Autopoiesis in the Game of Life. *Artificial Life*, 21(1):1–19.
- Buhrmann, T., Di Paolo, E. A., and Barandiaran, X. (2013). A Dynamical Systems Account of Sensorimotor Contingencies. *Frontiers in Psychology*, 4.
- Chalmers, D. J. (1997). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, New York, revised ed. edition edition.
- Clark, A. and Chalmers, D. (1998). The extended mind. *Analysis*, 58(1):7–19.
- Dennett, D. C. (1992). *Consciousness Explained*. Back Bay Books, Boston, 1 edition edition.
- Di Paolo, E. A. (2005). Autopoiesis, Adaptivity, Teleology, Agency. *Phenomenology and the Cognitive Sciences*, 4(4):429–452.
- Di Paolo, E. A. (2009). Extended Life. *Topoi*, 28(1):9–21.
- Di Paolo, E. A., Buhrmann, T., and Barandiaran, X. (2017). *Sensorimotor Life: An enactive proposal*. Oxford University Press, 1st edition edition.
- Egbert, M. and Cañamero, L. (2014). Habit-Based Regulation of Essential Variables. In Sayama, H., Rieffel, J., Risi, S., Dourzat, R., and Lipson, H., editors, *Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 168–175, New York, NY. MIT Press.
- Egbert, M. D. and Barandiaran, X. E. (2014). Modeling habits as self-sustaining patterns of sensorimotor behavior. *Frontiers in Human Neuroscience*, 8:590.
- Egbert, M. D., Barandiaran, X. E., and Di Paolo, E. A. (2010). A Minimal Model of Metabolism-Based Chemotaxis. *PLoS Comput Biol*, 6(12):e1001004.
- Egbert, M. D., Di Paolo, E. A., and Barandiaran, X. E. (2009). Chemo-ethology of an Adaptive Protocell: Sensorless sensitivity to implicit viability conditions. In Kampis, G., Karsai, I., and Szathmáry, E., editors, *Advances in Artificial Life, Proceedings of the 10th European Conference on Artificial Life, ECAL*, pages 242–250, Berlin. Springer.
- Hutto, D. D. and Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. The MIT Press, Cambridge, Massachusetts, 1 edition edition.
- Ruiz-Mirazo, K. and Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life*, 10(3):235–259.
- Stewart, J., Gapenne, O., and Di Paolo, E. A. (2010). *Enaction: Toward a New Paradigm for Cognitive Science*. MIT Press.
- Thompson, E. (2007). *Mind in Life*. Harvard University Press, London, England.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5:42.
- Varela, F. J. (1979). *Principles of biological autonomy*. North Holland New York.
- Varela, F. J., Thompson, E. T., and Rosch, E. (1992). *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press, Cambridge, Mass., revised ed. edition edition.
- Virgo, N., Egbert, M. D., and Froese, T. (2009). The role of the spatial boundary in autopoiesis. In *Advances in Artificial Life, Proceedings of the 10th European Conference on Artificial Life, ECAL*, pages 234–242, Berlin. Springer.