Extending Embodiment: Beyond Sensory-Motor Coordination

Francis Jeanson - candidate #64864
University of Sussex

Spring 2008

Abstract

This paper examines the particular theoretical issue of 'prescribed system design' in the embodied and situated systems approach for adaptive behaviour. It then offers an alternative perspective founded on Ashby's original framework.

Introduction

There is a growing ambition in the field of holistic artificial intelligence which claims that assimilating enaction in a theory of cognition should endow a system with genuine intelligent capabilities in the sense of adaptation to the environment for survival. There is from this motivation a particularly acute sense in which intelligent behaviour may only take place within a living organism. To formalize this condition, the autopoietic paradigm is held as the most profound and accurate (Maturana and Valera 1980). Although not all theorists depart from this conception, the general appeal to applying natural principles of living organisms to theorizing and modeling has gained significant ground in the artificial intelligence arena, and has often (but not exclusively) come under the label Artificial Life. Although many acknowledge that intelligence comes with life the conceptions remain disunited, the aims disperse and advancements sporadic. These advancements are not all meagre, but do actually constitute a fundamental base of principles that I will develop in a moment. Interestingly, modern efforts have begun to depart from the stricter sentence set out by the autopoietic view while embracing the direction towards which it points: the importance for an agent to have a 'way of life' (DiPaolo 2003). These views depart in that they claim genuine life as specifically prescribed by autopoiesis is not necessary for intelligent agency, but that intentionality on the behalf of the agent should suffice as minimal criterion.

Here I intend to argue against this view on pragmatic grounds. My aim is to defend the view that the mechanisms involved in the design of intelligent adapted behaviour cannot give rise to the richness and complexity exhibited by animals and humans without gaining itself and its constituents at some stage autopoietic complexity, therefor rendering the notion of 'true intentionality' vacuous. I will begin by exposing contemporary motivations regarding self-preservation as the sufficient condition for complex adaptive behaviour. I will then present how an alternate to this view which favours the notion of intentionality has offered to progress in the field; and how it may tie in with the valuable principles that have been learned in recent decades. This will expose specific weaknesses, which I hope to rectify by motivating a new principle: the requirement for 'trophism'.

Evading Internalism in Cognitive Robotics

Action without reason is implausible when attempting to understand the causes of behaviour. When observing an agent such causes are more easily attributed to external factors - such as the presence of a predator to explain the flocking behaviour in sheep. When no clear external factor can be found, an internal mechanism must be held responsible - such as muscular and nervous fatigue to explain sleep. In cognitive science however, a significant amount of work aims to explain all behaviour exclusively via internal processes. For instance, connectionist approaches seek to derive internal models of neural network structures to account for a range of cognitive phenomenon ultimately impacting behaviour. Although less obvious internalism is also deeply seated in the embodied and situated paradigm, even though coupling of an agent via a sensory motor apparatus is meant to rectify some disembodied pitfalls. The paradigm remains internalist in that a default behaviour is always explicitly defined for an agent by the designer or the criteria of selection in evolutionary scenarios. While it may be necessary to move 'inwards' in order to develop a powerful account of complex behaviour, the isolation of factors involved in behaviour to the internal actually seems to impose explanatory limitations: it appears increasingly difficult to explain behaviour beyond the reactive, the conditioned, or the embodied, without considering the impact of forces beyond sensory motor activity. Although hugely insightful and initially productive, the failure of the sophisticated and heavily funded COGS project at MIT (Brooks et al. 1999) to build a fully autonomous intelligent agent by the late 90's stands as evidence of an important misconception amongst the embodied/situated view of cognition. Further criticisms from the developmental viewpoint of ontogenesis are inline with this concern. Esther Thelen rejects the interactionist and transactionalist approaches on the basis of committing a logical fallacy by assigning a pre-existing plan in the form of internal instructions for the development of an organism (Thelen 2005). Hence establishing proper theoretical design principles of behaviourally sophisticated agents remains the principle challenge to date. I agree with DiPaolo and others that to overcome these design misconceptions it is necessary or at least useful to frame the reasons for actions with a fundamental principle: survival. DiPaolo, argues that an agent's active pursuit of self preservation renders actions that aid this self preservation intrinsically good, and those that go against this preservation intrinsically bad (DiPaolo 2003). Indeed acknowledging this reveals that any agent implementing this principle will act for its own account, thus presumably evading the internalist trap. The notion of survival stated simply however, does not suffice in my opinion to reveal this escape. The reality is that this notion of survival has been tackled in two distinct and opposing ways while neither attend to the problem of design in an adequate manner. The first is to dismiss the need to settle criteria for survival on the basis that agents will ultimately acquire the ability to self preserve with increased complexity in design. The second is to effectively (but not always explicitly) 'move away' from the problem of survival by reframing it in terms of intentionality. While the first view falls short for being a strict internalist approaches and fails for the same reasons as mentioned above, the second view however deserves careful attention. To do so, it is necessary to first examine important concepts brought forth regarding the matter of self preservation and adaptation.

Some scientific and philosophical developments from the past century can particularly aid in detecting the crucial mechanisms involved in adaptation and self preservation which have been recently motivating new conceptions of agent design. The most notable are scientific notions stemming from the cybernetic era, and philosophical developments from phenomenology.

Towards Genuine Self Preservation

The revival of cybernetic ideals in recent years has stimulated a vast number of theoretical and experimental prospects. We can think of the pioneering work by Grey Walter (1950) with his robot tortoises as a precursor to behavioural based robotics. Notably, original work on systems theory by Ross Ashby (1956) has been particularly influential in theories of adaptive behaviour with a systems theoretic approach. In his work he characterizes a number of fundamental criteria in order to better define the properties of complex adaptive systems. Amongst these, the notion of essential variables is of particular import. This notion is used to identify a set of key variables belonging to an agent that must be kept within certain bounds so as to maintain life. We can think for example of the heart rate which for humans must remain at a pulse rate roughly between 20 bpm and 300 bpm before serious nervous damage due to lack of oxygen or heart failure ensues. For systems and organisms in which variability is significant it is important not to confuse useful variables from essential ones. Yet this distinction in the literature has been mostly underdeveloped. Only those variables which are necessary for the maintenance of the agent's activity in *normal conditions* can be held as essential. These normal conditions may refer to either internal (metabolic) or external (environmental) states. A useful variable may be the fluctuation of light on the retina, which allows animals to avoid dangerous obstacles such as an on coming cliff. But cliffs are not part of the normal conditions of an animals environment, hence as long as the lack of light variation on the retina does not threaten the animals life in normal conditions, this type of variable is not essential. From this distinction then it is accurate to state that both internal and external states that constitute normal conditions are states which confer stability to the agent and to its coupling with the environment.

Due to laws of thermodynamics the environment and metabolism fluctuate continuously, causing an agent to be confronted with non normal conditions that perturb this stability. In order to deal with minor instabilities an agent may implement a behaviour-generating subsystem that couples via a closed sensory motor loop the agent with the environment, thereby regulating external variation via internal variation. These types of behaviour are typically considered as reactive, however a range of complex behaviours may still emerge from such basic coupling (Braintenberg 1984, Scheier et al. 1998). In the context of environmental adaptation this notion of sensory motor coupling may seem trivial, but it has in fact not been popular amongst the symbolic AI community, and has only been rediscovered since the cybernetics era by people such as Brooks, Cliff, Pfeifer and other proponents of embodied AI (Cliff et al. 1994, Pfeifer 1996). Scheier et al. as well as Nolfi have made important developments regarding models of behavioural control implementing tight sensory motor loops, showing how simple Khepera robots are able to discriminate via active perception objects that could otherwise not be discriminated from the environment via passive sensing alone (Nolfi 1997). Pfeifer however not only suggests that a closed sensory motor loop allows the agent to reduce the complexity of the perceived environment but that it also permits the formation of cross modal associations, thus making it one of the most fundamental principles of intelligent behaviour (Pfeifer 1996). As he notes, the ability for an agent to correlate information across a number of modalities is key for the agent's development of grounded concepts.

The implication that sensory motor coordination is responsible for the acquisition of meaning as Pfeifer suggests will be reiterated later in this discussion with respect to intentional behaviour. What is important to keep in mind is that because sensory motor coupling plays a foundational role in higher order functions, the mechanisms that instantiate it cannot be causally separated from the mechanism that implement these higher level functions.

If the environment or metabolism fluctuates severely a system's essential variables may reach life threatening values. Unless this severe fluctuations can be handled by a reactive response from the behaviour-generating sub-sustem a novel behavioural solution will have to be found. If this is the case an adaptive system must implement a mechanism that will control the behaviour-generating sub-system's parameters so as to effectively modify the system's behaviour until the essential variables regain equilibrium. For example a severe depletion of water in an animal's organism may trigger it to alter its current behavioural state and cause it to begin actively seeking water to ingest. According to Ashby, the mechanism that performs this (call it S), controls parameters of the behavioural regulatory system (call it R) and fulfills the capacity that self preserving systems must meet. This capacity is what Ashby calls ultra stability: a supplemental regulatory layer which monitors the status of essential variables and will introduce changes to the reactive system (the first regulatory layer) if these essential variables are threatened. This theoretical account of adaptive behaviour is particularly appealing in that it not only explains the origin of complex behaviour but also begins to account for certain fundamental cognitive ascriptions to intelligent agents such as motivations, desires, and intentions. Since the agent must maintain its essential variables within bounds to continue its existence, a mechanism that is driven by these essential variables to affect the parameters of the reactive system is thus self-driven, completing a causal cycle. This causal loop is completed because these essential variables are themselves affected by the environment or by the metabolism. Hence Ashby's framework successfully escapes the internalist trap. It remains incomplete however in that it doesn't sketch out any specifics for implementation, and fails to answer the following questions: What are a system's essential variables? How do these variables affect mechanism S, and how does S adjust the parameters of R? The two latter questions are currently being addressed by work in complexity theory, and system dynamics where principles of self-organization are beginning to reveal such mechanisms (Kelso 1995, Thelen 2005). The first question, on what these essential variables might be, constitutes the heart of the present debate. I gave an example earlier of heart rate for human beings, many other such examples are possible for a whole variety of organisms, but enumerating these is pointless as any system may potentially conform to new nonenunciated ones. Hence we need to establish what the defining criteria of essential variables actually are for any given system. It is crucial to have a clear grasp of the criteria that would allow us to design adaptive agents without falling prey of the internalist trap. Two possible view may bare fruit at this stage: the first is to favour a teleologically inclined form of reasoning about what constitutes essential variables for an adaptive agent, or the second more fundamental appeal which characterizes essential variables in virtue of life promoting dynamical organization. In the following I will exhibit the first, and attempt to show the limitations it imposes on the field of design for cognitive control systems. I then defend the second as the adequate take for the successful understanding and design of intelligent behaviour.

The Escape via Intentional Behaviour

In a 2003 paper, DiPaolo argues in favour of the phenomenological approach to agency which defends the view that perception and action cannot account alone for the circumstances of intelligent behaviour. Instead, effective agency must be complemented with genuine intentionality (Merleau-Ponty 1963). By doing so, it is claimed that an intrinsic form of meaningful existence will be enjoyed by the agent. Under such conditions, it therefor seems possible for an agent to acquire a genuine sense of life necessary for its self preservation and to satisfy the essential conditions as defined by Ashby. Although the phenomenological insight of having to go beyond the sensory and motor is essentially correct (as I defend in the previous section), talk of intentions appears to qualify the agent with purposeful existence a priori. Acknowledging that an agent is *driven* by purpose is deductively correct but fails to ontologically ground the very process which leads to active subsistence. Indeed, as DiPaolo points out, we may be tempted to interpret the sense of meaning as it is employed in phenomenology as the traditional sense of semantic meaning, which is wrong. As mentioned above, the principle of sensory motor coordination can effectively account for the semantics to be picked out from the agent's interaction with the environment and solve the famous symbol grounding problem (Harnad 1990, Scheier and Pfeifer 1995, Pfeifer 1996). But the phenomenological sense of meaning appeals to the prospect of awarding genuine purpose to the agent, but how does that excuse it from a grounding of its own? The fact is that it does not. There is no clear sense in which purpose has any significance in a context void of any agency. Does wind serve purpose for the rock? Although it may shape its form, smoothen its surface or even displace it with force, there seems to be no coherence worth extracting from such a notion, because purpose and utility are only relevant if the target is a living agent. Intentionality may be a product of genuine agency but it cannot serve as grounding for it. Hence the teleological approach for the characterization of essential variables does not fit the bill. We cannot resolve the internalist trap by overarching the requirements for adaptive agency by prescribing artificial purpose, this is precisely the trap we need to evade.

In order to complement his view however, DiPaolo develops the notion of habit formation. According to him a framework that favours a complete account of living systems cannot, at least at this state, offer a way to explain specific instances of behaviour. Yet because intelligent behaviour is largely underdetermined, an effective approach must offer a coherent view for why agents may adopt a specific strategy rather than another to achieve a particular task. There is no doubt that the problem of behavioural underdeterminism must be resolved (Bernstein 1967). It is also true that this problem bares significance for an adequate account of adaptive behaviour. Models of adaptive behaviour, such as DiPaolo's habit formation proposal, will have to offer valuable explanatory and predictive potential. In short form, this proposal suggests that an agent's behavioural habits are invariants (alike Ashby's essential variables) that a system maintains via circular causal coupling between mechanisms for plasticity and behaviour, where behaviour induces plasticity and plasticity modifies behaviour. Indeed DiPaolo may be offering a solution to the underdeterministic problem of behaviour, but this isn't the present concern. What he claims, is that by adopting this proposal an adequate level of modeling can be reached to climb the 'complexity ladder' of sophisticated adaptive behaviour without resorting to the viability constraints which are favoured by Ashby, Maturana and Valera, and the like. Although I do not disagree that the process of habit formation as stated by DiPaolo may serve purpose in refining current conceptions of adaptive behaviour, It seems dangerous to acknowledge that such a proposal wouldn't falter into internalism. The strength of his argument stems from the ascription of a system's invariant to the notion of habit. By doing so it seems paussible to skip the viability criteria of essential variables required in Ashby's framework. However by closing the causal loop from behaviour to plasticity, this view fails to recognize the weakness of the operational domain in which it applies: behaviours induce plasticity in virtue of external environmental resources, and plasticity modifies behaviour in virtue of internal resources. Yet the energy necessary to drive the internal mechanism cannot be independent from the external, or it would violate laws of thermodynamics. This would not be problematic for this view if it did not commit to serve as the *basis* for explanation of agent adaptive behaviour. The proposal of habit formation as a fundamental principle for adaptive behaviour may prove to be highly valuable, but alone it cannot justify the intrinsic drive, nor the level of complexity of sophisticated intelligent agents.

To do so I believe it is important to appreciate the potential of recent intuitions regarding the basis of intelligent agency: the requisite for trophism.

Requisite Trophism

There is a strict sense in which the notion of life implies fundamental principles of material organization. According to the autopoietic view developed by Maturana and Valera, living organisms differ from the non living in that they are capable of self-renewal, self-maintenance and stability within a particular milieu (Maturana and Valera 1980). The organization of an autopoietic system is the result of such self-renewal and sell-maintenance in virtue of a network of component producing processes such that interactions between these components gives rise to the very same network of processes that produced them. Hence a circular mechanism is in place. In this sense the organization of an autopoietic system is closed. Archetypical examples of this sort of system are cells, organs etc... In contrast, allopoietic systems are those which are organizationally open. Here organization is the result of networks of processes that also produce components, but these components do not in turn give rise to a self-maintaining network. Such as crystalline structures, molecular chains, etc... The present concern is not to debate a proper definition of life, what matters in this context is the value this view can bring to the problem of adaptive behaviour. An autopoietic system is a physically bound organization, in which catalytic-like processes take place to reproduce and maintain the system's organization. However because all systems abide by the laws of thermodynamics and are placed in a some given environment, these organizations are subjected to external interferences and disruptions. Hence we can conceive of these minimally self-organizing systems as potentially adaptive if they are to subsist. To frame an adaptive system as autopoietic forces the view that adaptation does not necessarily come hand in hand with sophisticated macroscale organisms. A unicellular organism such as the bacterium Escherichia coli may dispose of a wide range of internal components to maintain its internal organization, but also to sense the presence of food and move about its environment using pseudo-pod (Wolpert et al. 2002). What is important to see at this microscale is that any system with organizational closure depends on the environment as source of energy so that it may pursue this self-generation, self-maintenance and stability. This precise requirement is a necessary condition to solve the problem of internalist design, and precisely characterize what constitutes the essence of essential variables. What is intended by the notion of trophism relates specifically to requirement that any self-organising system can only come into being via the presence of preexisting energetic conditions so that the necessary reactions for organization can occur and subsist. The term trophism is borrowed from the biological sense used to distinguish heterotrophic from autotrophic organisms.

This condition however raises a number of issues, for which only a few can begin to be answered. How does this condition imply that current robotics are not constructing energy absorbing robots? Can simulations of autopoietic systems provide the backdrop for this condition? How does this condition help the design of behaviourally more sophisticated agents than that which is possible today? The first can be answered relatively succinctly. Robots that absorb energy *are* escaping the internalist trap, but only in the theoretical conception emphasized here. There is a sense in which the lack of energy absorbed by 'wall socket searching', solar, or wind robots does not constitute a threat to their integrity: a robot simply switches off. Hence the energy is not a quantity dictating the essential variables in Ashby's sense. We may still be tempted to say that the robot dies, but if it is not organized in the autopoietic sense then it was not living in the first place. If it was organized in the autopoietic sense then the lack of energy would inevitably constitute a disruption apt of threatening the agent's essential variables. This is why the autopoietic conception is intimately tied to the requirement for trophism. But alone is does not stipulate the conditions of required energy detection, absorption and consumption. The second question is ill conceived, because the energy necessary to simulate a trophic autopoietic system is also artificial. What we want however is a formal understanding of the principles of complex adaptive behaviour, hence we do not need to reproduce life as it is, but only need to focus on functionally equivalent mechanisms that could be implemented in design. The more difficult problem seems to be at the physical hardware level. The third question requires extensive development, and I can only answer partly. By following Ashby's guideline and taking into account the mechanisms involved in genuine self preservation as proposed here, efforts in systems and complexity theory should lead to important new insights. Turing's work on the chemical basis of morphogenesis has inspired work in neural dynamics of mass action (Turing 1952, Katchalsky 1971, Freeman 1975). Yet these same dynamics have become foundational in the ontogenetic development arena (Wolpert et al. 2002, Ikegami and Suzuki 2007). Although these dynamics tend to overcast the importance of integrating an energetic term with abstract algebraic terms. The sort of descriptions achieved via dynamical systems theory, should reach the status of proper causal accounts in virtue of this requisite. Hence the design of robots that do not simply behaviourally mimick but that causally engage in their environment should be increasingly possible.

Conclusion

This paper attempted to motivate the gaining concern that, as is, embodied robotics is not enough to account and achieve the design of agents which display complex adaptive behaviour. This approach succumbs to what has been denoted as the 'internalist trap': that designed systems are predisposed with agency, and lack genuine criteria for viability because of it. It was shown that original work from the cybernetics age of the 50's has made important contributions to address the issue. Extending these, modern developments have typically either favoured a position that claims life if necessary to fit the criteria of maintained internal stability, whereas others prefer an intentional take which skips the difficult barrier in attempting to generate living systems in the autopoietic sense. Weaknesses of the latter view have been exposed in favour of the first. Finally, a contribution is made which suggests that by factoring in the influence of energy on a system's physical organization within a dynamical framework, should help guide advancements in adaptive systems design in fruitful ways.

References

Ashby, W. R. (1956). An Introduction to Cybernetics, (Chapman & Hall, London).

Ashby, W. R. (1960). Design for a brain: The origin of adaptive behaviour (Second edition). London: Chapman and Hall.

Bernstein, N. A. (1967) The Coordination and Regulation of Movements, Pergamon, London.

Braitenberg, V. (1984). Vehicles: Experiments in synthetic psychology. Cambridge, MA: MIT Press.

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., & Williamson, M. (1999). The Cog project: Building a humanoid robot. In Nehaniv, C. (Ed.), Computation for metaphors, analogy, and agents, Lecture Notes in Artificial Intel ligence 1562, pp. 5287. New York, Springer-Verlag.

Cliff, D., Husbands, P., Meyer, J. A., Wilson, W. (1994). From animals to animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior. MIT Press. p.91-99

Di Paolo, E. A. (2003). Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In K. Murase and T. Asakura (Eds.) Dynamical Systems Approach to Embodiment and Sociality (pp. 19-42). Adelaide: Advanced Knowledge International.

Ikegami, T., Suzuki, K. (2007) From homestatic to a homeodynamic self. BioSystems 91 388-400.

Freeman, W. (1975) Mass Action in the Nervous System. New York, Academic Press.

Grey, W. W. (1950) "An Imitation of Life," Scientific American, May 1950, p42-45.

Harnad, S. (1990) The Symbol Grounding Problem Physica D 42: 335-346.

Katchalsky, A. (1971) Biological flow structures and their relation to chemodiffusional coupling. Neuroscience Research Program Bulletin. 9, 397-413.

Kelso, S. (1995) Dynamic Patterns: The Self-Organization of Brain and Behavior. Bradford Books. MIT Press.

Maturana, H., Varela, F. J. (1980). Autopoiesis and cognition: The realization of the living. Dordrecht, Holland: D. Reidel Publishing.

Merleau-Ponty, M. (1963). The structure of behaviour. London: Methuen. Translated by A. L. Fisher.

Nolfi, S.(1997) Evolving non-trivial behaviors on real robots: A garbage collecting robot. Robotics and Autonomous System, 22, pages 187–198.

Pfeifer, R. (1996) Building Fungus Eaters: Design Principles of Autonomous Agents, Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour, MIT Press, Cambridge MA.

Scheier, C., Pfeifer, R. (1995). Classification as sensory-motor coordination: A case study on autonomous agents. In Moran, F., Moreno, A., Merelo, J. J., and Chacon, P. (Eds.), Proceedings of the 3rd European Conference on Artificial Life, pp. 657 667 Granada, Spain. Springer Verlag, Berlin.

Scheier, C. Pfeifer, R., Kunyioshi, Y. (1998) Embedded neural network Exploiting constraints. Neural Networks, (7-8):1551-1569.

Thelen, E. (2005) Self-Organization in Developmental Processes: Can Systems Approaches Work? In Brain Dovelopement and Cognition: A Reader, Mark H. Johnson, Yuko Munakata, Rick O. Gilmore, 336-374

Turing, A. M. (1952)The Chemical Basis of Morphogenesis. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 237, No. 641, 37-72

Wolpert, L., Beddington, R., Jessell, T., Lawrence, P., Meyerowitz, E., Smith, J., (2002) Principles of Development, Second Edition Oxford University Press.