Workshop on Artificial Autonomy ECAL 2011
Twenty Years of Practice of Autonomous Systems
August 8, 2011
Paris

Book of Abstracts

Autonomy

Autonomy is a system's capacity to produce, sustain and regulate itself and its relationship with the environment it co-determines. Ultimately autonomous systems are those capable of defining them-selves (autos) and their norms or rules of operations (nomos).

In the Proceedings of the First European Conference on Artificial Life [Varela & Bourgine (1991) Towards a Practice of Autonomous Systems. Proc. of the First ECAL, pages xi-xvi] a full research program on the artificial implementation of autonomous systems was envisioned. According to that view, autonomy was the fundamental property underlying both the phenomenon of life and cognition, so the challenge of the new artificial sciences (if they were to illuminate our understanding of those phenomena) consisted in the simulation or realization of systems with increasingly autonomous capacities. Artificial autonomy was taken up as a difficult but achievable research goal, the pursual of which could contribute to the merging of paradigms, the new, Artificial Life and the old, Artificial Intelligence, and moreover, could help to uncover some of the most elusive properties of biological organization.

The main topic of the workshop is to evaluate the 20 years of practice of autonomous systems from that first ECAL up to the ECAL taking place this August, 2011; to review the state of the art and the future challenges and research milestones. The notions of autonomy, autopoiesis, emergence, enaction, etc. have played a central role in Artificial Life but have also experienced a relative displacement as the field has made progress on more specific and specialized research fields (from network topology to ant intelligence, from evolutionary robotics to multicellular development). We believe that increasingly specialized trends in Artificial Life should also be balanced with renewed attempts for integrative approaches. In this sense, reviewing and recovering the visionary and pioneering notion of autonomy 20 years later provides a challenging opportunity to re-assemble Artificial Life back into the organism-centered, emergent, holistic yet mathematically and conceptually mature notion of autonomy.

Special Challenge

How to model mind in life and life in mind? Much progress has been made in two directions, biological autonomy (protocell research) and behavioural autonomy (cognitive/adaptive robotics), but integrating the progress made in these two areas remains an open challenge. How can we model an integrated autonomous system capable of displaying both self-production and flexible, self-regulating behaviour?

Table of Contents

On the chemical relevance of compartment 'self-production'

Ben Shirt-Ediss[1,2,*]
Ricard Solé[1,3]
Fabio Mavelli[4]
Kepa Ruiz-Mirazo[2,5]

[1] ICREA-Complex Systems Lab, Universitat Pompeu Fabra, Spain
[2] Department for Logic and Philosophy of Science, University of the Basque Country, Spain
[3] Santa Fe Institute, USA
[4] Department of Chemistry, University of Bari, Italy
[5] Biophysics Unit (CSIC-UPV/EHU), University of the Basque Country, Spain
[*] Corresponding author: ben@shirt-ediss.me

Summary

We propose an experiment to systematically investigate what relevance a membrane has to the dynamics of an enclosed chemical reaction. We are particularly interested in the case where the membrane, in particular the bi-layer lipid membrane of a vesicle, not only encloses, but is also produced by the internal reaction. Three different membrane boundary cases are proposed for investigation, and four reaction schemes to initially explore in each case are detailed. Results of the study, which is to be carried out as a stochastic computer simulation, are to be presented at the ECAL 2011 Workshop on Artificial Autonomy.

Extended Abstract

An interesting class of chemical reaction system is one in which the boundary of the system is also included as part of the description of the chemical dynamics. This is in fact necessary in order to achieve a correct understanding of protocell systems, as they are being currently investigated (Hanczyc et al. 2007; Mansy et al. 2008; Rasmussen et al. 2009; Maurer et al. 2009; Budin & Szostak 2011), because the different chemistries implemented are highly dependent on the type of lipid compartments involved.

By building and/or influencing the properties of a self-enclosing membrane, a reaction system exists in its own "micro-environment", where the internal chemical dynamics can have direct effects, for example, on the elasticity of the boundary, or on its permeability to food resources and waste compounds. Able to modify its own conditions of stability (viability range) or to selectively change its coupling to the local environment from within, like all cellular organisms actually do (but in much more sophisticated ways, see Mouritsen 2005), this class of "membrane-metabolism" reaction systems (self-producing vesicles or, more simply, 'protocells') is particularly relevant to the study of minimally autonomous agents. Indeed, if appropriately organised, one can imagine a scenario where the membrane-metabolism coupling is such that the protocell system manages to effectively "channel" energy and matter through itself and persist in a far-from thermodynamic equilibrium state, in the face of significant environmental perturbation (Mavelli & Ruiz-Mirazo 2007: 2008; Ruiz-Mirazo & Moreno 2004).

Using stochastic simulation, the goal of this work is to clarify more precisely how the dynamics of a simple chemical reaction network play out differently in the following boundary scenarios:

B1. Reaction network with no boundary. The reactions proceed in bulk homogeneous aqueous solution, confined just by the inert walls and geometry of the reactor.

B2. Reaction network which influences a lipid bi-layer boundary. The reactions proceed within

a pre-given self-assembled lipid vesicle micro-environment. However, the chemical network cannot directly change the composition of the membrane, but only have some influence on it, for example, by producing osmotic tension.

B3. Reaction network which both produces and influences a lipid bi-layer boundary. As in case 2 above, but now the chemistry also involves the synthesis of lipids which become incorporated into the pre-existing membrane, modifying its composition, or self-assembling into newly formed vesicle membranes.

The last case B3 is the one of higher interest, but the previous two will serve as control experiments. Figure 1 gives a graphical illustration of the scenarios. The four simple chemical reaction sets (R1-R4) detailed below will be analysed in case B3 (where the reaction set can produce and influence the boundary). For the control experiments, the same reaction sets (R1-R4) will be used, but this time substituting a generic compound Q for lipid L. Unlike L, compound Q is not amphiphilic and it cannot form or incorporate into the vesicle membrane.

Candidate Reactions

The initial reaction is just a reaction that produces lipid L and a waste molecule W from a resource R, with the help of a catalyst C:

$$R+C \leftrightarrow L+C+W \qquad (R1)$$

Next, the same reaction is considered, but now producing an extra catalyst molecule. The rate of production of products exponentially increases because the amount of available catalyst increases:

$$R+C \leftrightarrow L+2C+W \qquad (R2)$$
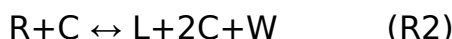
Thirdly, a Lotka-Volterra style reaction scheme is to be analysed, in which the production of the lipid L takes place by means of two metabolites (X and Y), in such a way that the latter spontaneously show oscillatory behaviour:

$$R+X \rightarrow 2X$$
$$X+Y \rightarrow 2Y$$
$$Y+C \rightarrow L+C+W \qquad (R3)$$
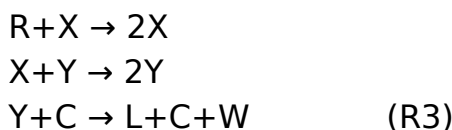
Finally, a minimal reaction which acts like a toggle switch (Wilhelm 2009), showing bi-stable behaviour for the production of lipid L will be investigated:

$$R+C \rightarrow 2L$$
$$2L \rightarrow L+C$$
$$L+C \rightarrow C+W$$
$$L \rightarrow W \qquad (R4)$$
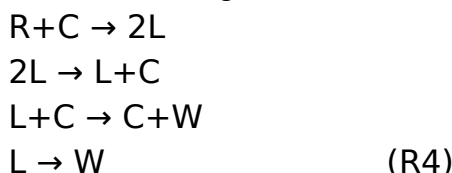
The main question we want to address, through analysis of the above reaction sets in the different boundary cases B1-B3, is the possible advantages in terms of the stability and robustness brought to a chemically reacting system when it is enclosed by a semi-permeable lipid membrane, particularly if this is of its own producing. Additionally, the synchronization of the chemical reaction dynamics (e.g., oscillations in the concentration of metabolites, like in reaction R3) with the cellular cycle (periodic division/reproductive events) is to be explored.

Experimental Procedure

Experiments proceed by first executing the candidate reaction in bulk, in a simulation of a Continuous Stirred Tank Reactor, and then introducing a single vesicle to encapsulate a proportion of the reactor contents (Figure 1). The outplay of the reaction dynamics in the initial vesicle, as well as the behaviour of the possible ensuing population of vesicles will be observed. The diameter of the initial vesicle is to be varied, as are the relative proportions of reactor contents initially trapped

inside it, which can vary quite largely (for example, see Luisi et al. 2010).

The robustness of dynamical behaviour observed in the system can be assessed by (a) peturbing the inflow of substrates and water into the flow reactor in a parameterised way, and (b) by changing the kinetic constants of the reaction networks and parameters of the lipid vesicle model (such as the permeability of the different solutes across the membrane), to see in which parameter ranges similar behaviour is observed.

Prospective Insights From This New Approach

The entrapment of a chemistry inside a micro-environment, as performed in this work, is interesting for a number of reasons.

In general, reactions taking place in a micro-volume (like the aqueous pool inside a vesicle, typically $5.2 \times 10\text{-}19$ litres), are subject to significant fluctuations of chemical species as, due to the exceedingly small volume, often only a handful of molecules are needed to cause large concentration differences. If the chemical reaction network is close to instability, this factor can change the dynamical outcome significantly [1] [2].

Secondly, we expect reactions to proceed with different dynamics when encapsulated in the aqueous pool of a lipid vesicle, because, due to the structure of the vesicle system, the chemical equations for the internal reaction network become additionally coupled through other relating factors.

Going back to the reaction tank scenario briefly (case B1), we can observe here that the chemical equations of the reaction system along with the inflow and outflow properties of the reactor determine the system dynamics[3]. The volume of the bulk can change, but this change will not have an appreciable effect on the reaction rate constants if the tank is large.

However, in the vesicle environment scenarios (cases B2 and B3), the concentration of the aqueous pool reactants determine the amount of water flow through the membrane (Donnan effect), changing the internal volume and influencing the reaction rates in a very significant way. The vesicle membrane adds another level of dynamical coupling, because the tension in, and surface area of the membrane determine the inflow and outflow of substrates and products for the internal reaction network. The internal reaction network can cause osmotic tension in the membrane by the Donnan effect mentioned above, or can produce lipid (case 3) to be incorporated within and grow the membrane.

Finally, the ability to grow the membrane (or, conversely, shrink the volume of the aqueous pool), grants one dynamical feature to the vesicular reaction, that its counterpart proceeding in bulk can never accomplish: division into daughter vesicle "micro-environments".

References

Budin, Itay, & Szostak, Jack W. 2011. Physical effects underlying the transition from primitive to modern cell membranes. PNAS, 108(13), 5249–5254.

Hanczyc, Martin M, Toyota, Taro, Ikegami, Takashi, Packard, Norman, & Sugawara, Tadashi. 2007. Fatty acid chemistry at the oil-water interface: self-propelled oil droplets. Journal of the American Chemical

---

1For this reason, the Stochastic Chemical Kinetics framework, dealing with discrete molecule numbers as opposed to concentrations (continuous values only properly defined for a very large population of molecules), is generally deemed best for modelling intra-cellular environments.

2In small volumes, molecular crowding, particularly prominent with the presence of large macro-molecules, is also a significant factor influencing the calculation of reaction rates, but this factor is not intended to be explored in this work.

3 Temperature and pressure of the reaction are also relevant factors, but these aspects are not covered in this study.

Society, 129(30), 9386–91.

Luisi, Pier Luigi, Allegretti, Matteo, Pereira de Souza, Tereza, Steiniger, Frank, Fahr, Alfred, & Stano, Pasquale. 2010. Spontaneous protein crowding in liposomes: a new vista for the origin of cellular metabolism. Chembiochem, 11(14), 1989–92.

Mansy, Sheref S, Schrum, Jason P, Krishnamurthy, Mathangi, Tobé, Sylvia, Treco, Douglas a, & Szostak, Jack W. 2008. Template-directed synthesis of a genetic polymer in a model protocell. Nature, 454, 122–5.

Maurer, S E, Deamer, D W, Boncella, J M, & Monnard, P-A. 2009. Chemical evolution of amphiphiles: glycerol monoacyl derivatives stabilize plausible prebiotic membranes. Astrobiology, 9(10), 979–87.

Mavelli, Fabio, & Ruiz-Mirazo, Kepa. 2007. Stochastic simulations of minimal self-reproducing cellular systems. Philosophical Transactions of the Royal Society B, 362, 1789–1802.

Mouritsen, Ole G. 2005. Life - As a Matter of Fat: The Emerging Science of Lipidomics. Springer.

Rasmussen, Bedau, Chen, Deamer, Krakauer, Packard, Stadler (eds). 2009. Protocells: Bridging Nonliving and Living Matter. MIT Press.

Ruiz-Mirazo, Kepa, & Mavelli, Fabio. 2008. On the way towards 'basic autonomous agents': Stochastic simulations of minimal lipid-peptide cells. BioSystems, 91, 374–387.

Ruiz-Mirazo, Kepa, & Moreno, Alvaro. 2004. Basic autonomy as a fundamental step in the synthesis of life. Artificial life, 10(3), 235–59.

Wilhelm, Thomas. 2009. The smallest chemical reaction system with bistability. BMC systems biology, 3, 90.

Affective agents bridge the gap between life and mind

Theije Visser *theijevisser@gmail.com*
Rineke Verbrugge *l.c.verbrugge@rug.nl*
Tjeerd Andringa *tjeerd@ai.rug.nl*

Department of Artificial Intelligence
University of Groningen
July 18, 2011

Summary

We develop an agent-based model which implements affect regulation as the motivation for decision-making in individual affective agents within a simulated social environment. The model design applies the life-mind continuity thesis to the field of social simulation. Agents are fully coupled to the environment through perception and action via core affect regulation. This makes affect the center of the regulation of the coupling between the agent and the environment. The model is used to generate the emergence of zones of cooperation in the Demographic Prisoner's Dilemma (DPD) paradigm. This model could provide a more biologically and cognitively plausible explanation for the dynamics of cooperation within the DPD. Although the model abstracts away from self-production and identity generation, it is sufficient to generate these dynamics and provides blocks to build more complex models of human behavior based on affect self-regulation and self-production to gain a deeper understanding of the behavioral foundations of the emergence of social dynamics.

Introduction

The *affective learner* model that we introduce in this paper integrates high-level interactional dynamics and low-level behavioral autonomy through an agent-based model which implements *affect regulation* [8] as the motivation for behavior. Current agent-based models within the field of social simulation research (see [10, 7]) focus chiefly on rule-based behavior, without taking into account the intrinsic coupling between the agent system and its environment, which is always present in autonomous agents [11]. This intrinsic coupling is defined in a theory of biological agency [4, 2]. We aim to enrich social simulation research by building the agents which constitute a simulated society as autonomous systems which display self-regulating behavior. We use the *life-mind continuity thesis* to build the bridge between agency and social dynamics. This thesis states that life and mind have a common set of basic organizational properties (see [9], [3]).
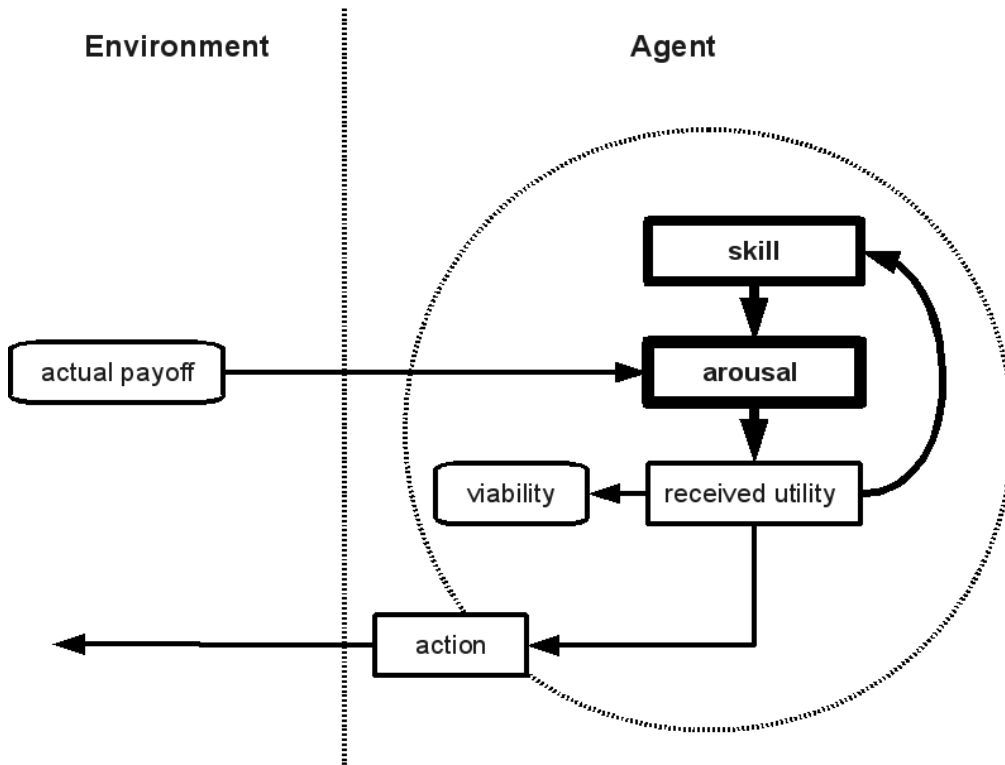
Figure 1: The relation among agent components in the affective learner agent. The social skill provides feedback to the viability value, through the choice of arousal.

We create a more biologically plausible model of behavior by designing agents which adhere to the conditions of biological agency. Our simulation of a society of affective agents combines insights from three scientific disciplines. First, a behavioral model based on the theory of biological agency [4, 5] provides the elementary agent design. Second, insights from affect research are used to operationalize agent behavior through affective motivation and self-regulation [12]. Finally, the interaction design is formalized by utilizing game theoretic tools [1, 6].

## *The affective learner agent*

Behavior of living beings is in the first place aimed at maintaining existence by keeping the individual (far) from death. Di Paolo [4] calls the distance from death *viability*. Maintaining viability is essential for all living entities. An explicit definition of *biological* agency was recently proposed [4, 2]. In this definition, an agent is defined as a self-constituting system that adaptively regulates its coupling with its environment and contributes to sustaining itself as a consequence. An agent acts by modulating the coupling between the agent state and the environment state.

Because our model is an attempt to implement a model based on complex components inspired by the artificial autonomy paradigm, we chose to abstract away from self-organization by externally imposing connections between the components. As stated, our model investigates the life-mind continuity thesis in the form of an affective learner. It is important to note here that we aim to inform the research on life (sociology) by building an individual-based model with insights from mind-research based on the theory of biological agency.

The affective learner agent regulates its relation to the environment to lead its state away from the limits of viability. Furthermore, because all living agents have finite resources for action, each identified course of action needs to be evaluated in terms of a cost-benefit analysis before it is initiated [12]

An affective learner agent exhibits three main components: viability, skill, and arousal. It reacts to its environment by predicting future affective states. This way, actions of an agent are based on action selection which should result in the most viable state possible.

ACTIVATION
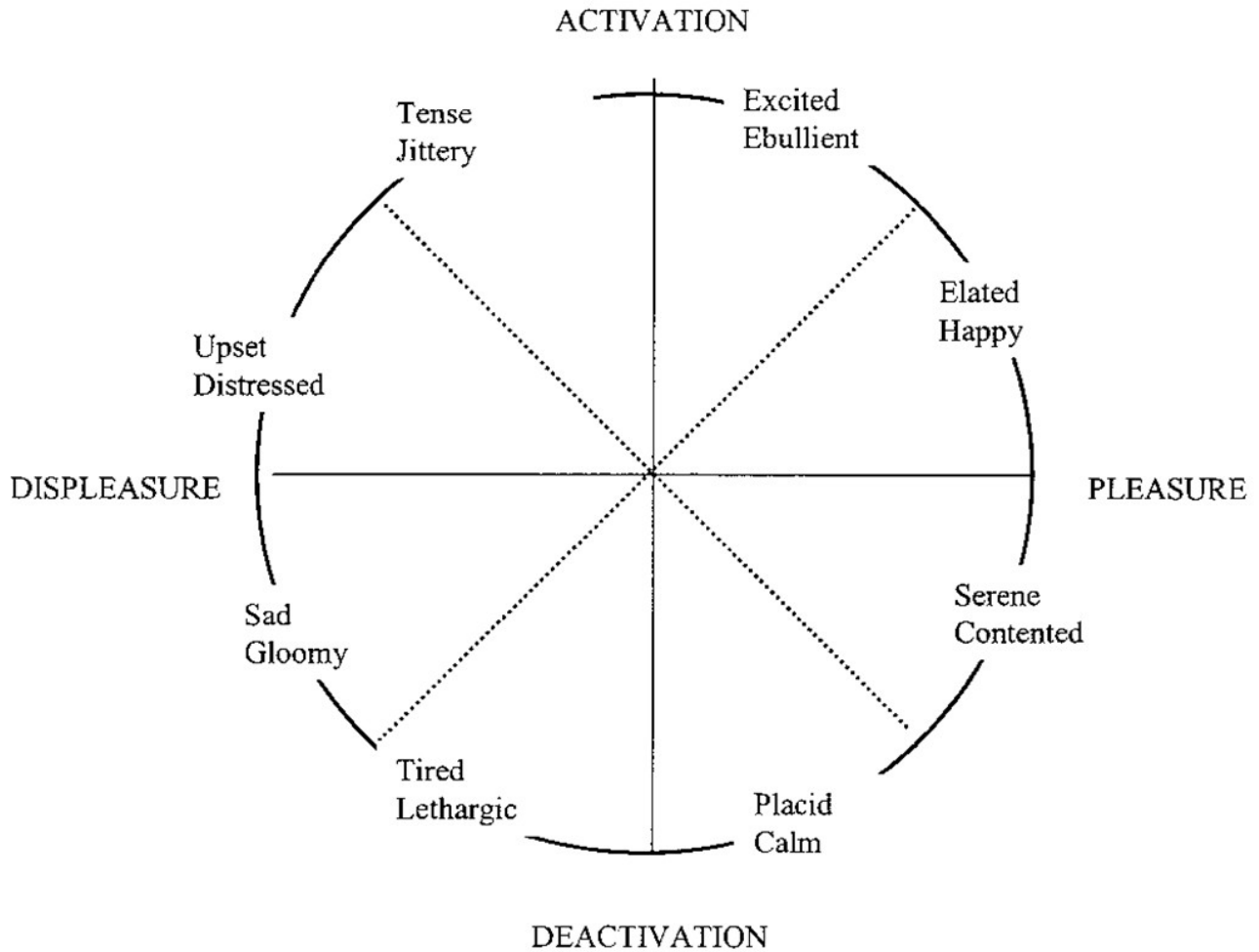
Figure 2: The two dimensions of core affect (reproduced from [12]).

The combination of two functional demands in the definition of agency, namely access to the agent's viability and resource allocation, have been identified in cognitive science as the notion of *core affect* [12], which consists of two dimensions: viability and arousal (see figure 2).

 The agent acts by first predicting which action will provide the optimal change in viability, by 'simulating' the action. When it has chosen an action, it makes the action 'hot', by investing a certain amount of affective resources, reflected by the arousal value. The social skill level determines the degree to which the agent makes action choices which result in optimal affective impacts. Skill learning depends only on the quality of perception of the expected affective impact of an action. It is learned by making mistakes: learning is dependent only on the magnitude of the mistake, not on the direction.

According to Di Paolo [4], to distinguish a genuine agent, three conditions have to be met: individuality, interactional asymmetry, and normativity. The affective learner agent does adhere to all three conditions of biological agency (see section 2). First, it implements the condition of interactional asymmetry, since it is the source of activity in the environment. Furthermore, it adheres to the condition of normativity, as it takes deliberate, intentional action according to a norm

based on its internal state. Finally, the model adheres to the individuality condition, by adding a motivational component, which modifies the affective influence of interactions through a learning mechanism. This way, the system is defined by the interconnectedness of all variables making up the network (see figure 1).

Social dynamics

To study a *society* of affective learner agents, we utilize the Demographic Prisoner's Dilemma (DPD) paradigm [6] to implement the interaction model. We propose that if the affective learner agent does replicate the results of some of the extensions of the DPD model, then the affective learner model can be seen as representing the low-level affective motivation of behavior underlying the DPD dynamics.
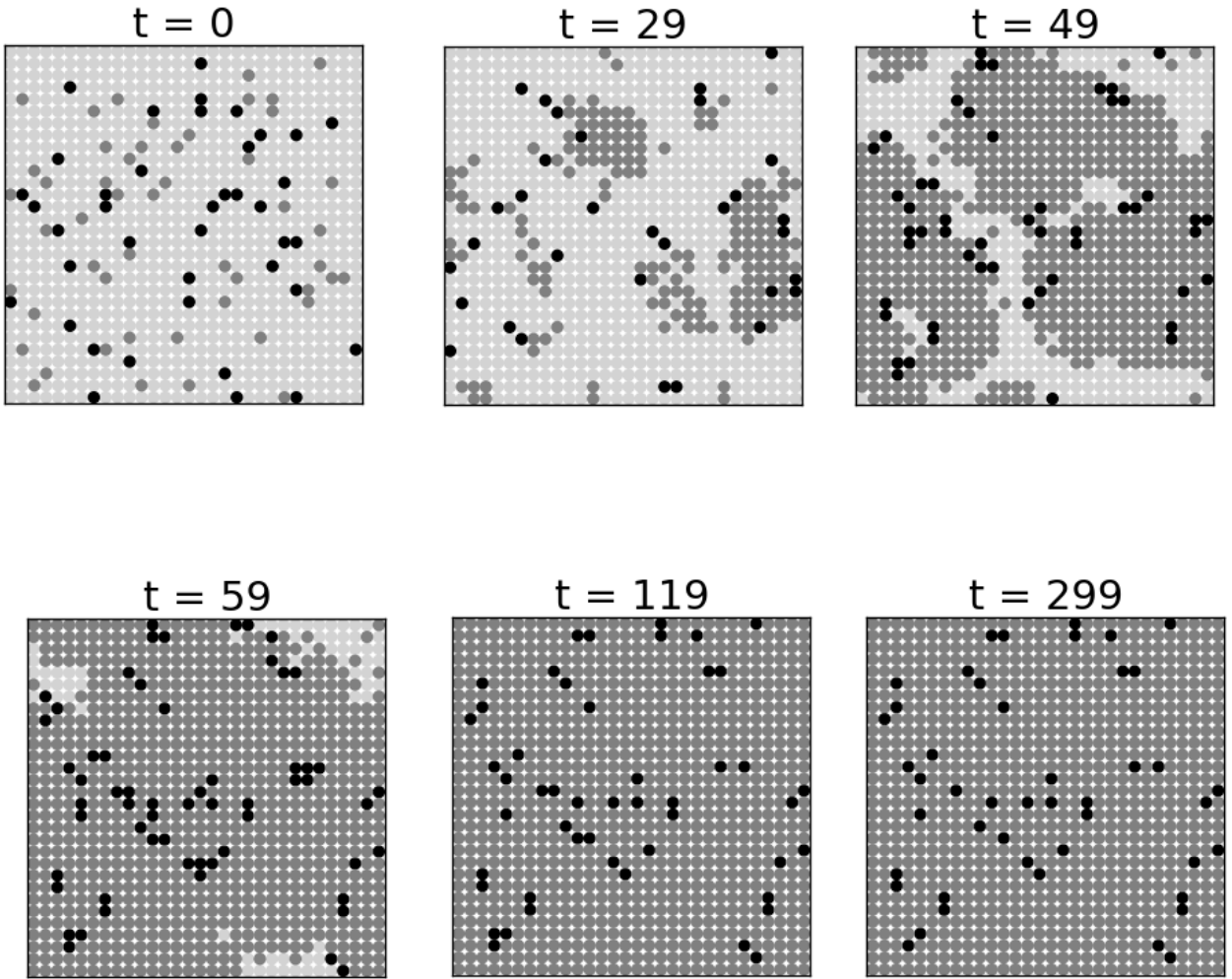
Figure 3: Snapshots of the spatial configuration at time-steps t for a run of the affective learner agent population over 300 steps.

Agents are randomly distributed on a 30x30 torus world. Each agent is initialized as one of the two fixed types: cooperator or defector. Every time-step, agents play a game with each neighbor within the Moore neighborhood. From every interaction, the agents receive a payoff according to the Prisoner's Dilemma payoff matrix [1, 6]. An experiment was performed on an environment in which agents can reproduce and die in a manner specified by the DPD paradigm (see [6], p. 206).

Figure 3 shows that first all cooperators gather into clusters, while defectors typically gather around these cooperator clusters. These mixed clusters are stable for some time, until too many defectors enter the cluster, and cooperators experience a negative viability change, after which they will move elsewhere.

The population dynamics as illustrated by the phase diagram (see figure 4, right panel) show that an equilibrium occurs only after each cell is occupied. First, we see that zones of cooperation form between $t=29$ and $t=49$. After this, an equilibrium forms around $t=80$ (see figure 4, left panel). When we look at the mean values (see figure 5), cooperators are able to optimize their viability level, while their skill level drops to 0.2 throughout the run. Defectors keep the mean viability level at around 0.5, while the skill level rises to 0.3 throughout the run.
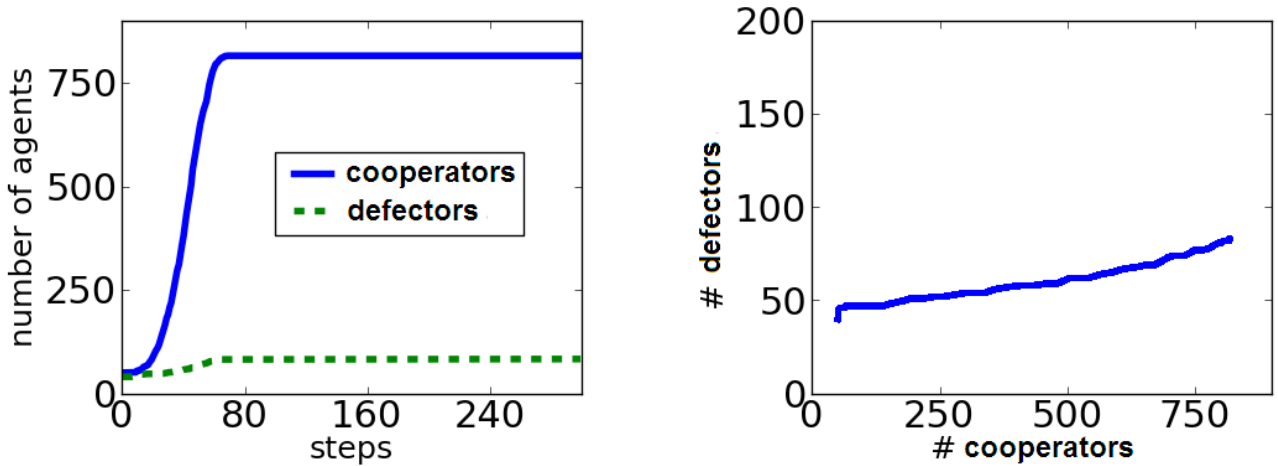
Figure 4: The left panel shows the population dynamics for a society of affective learner agents over 300 steps. The right panel shows the phase diagram, which represents the ratio between the number of cooperators and defectors.
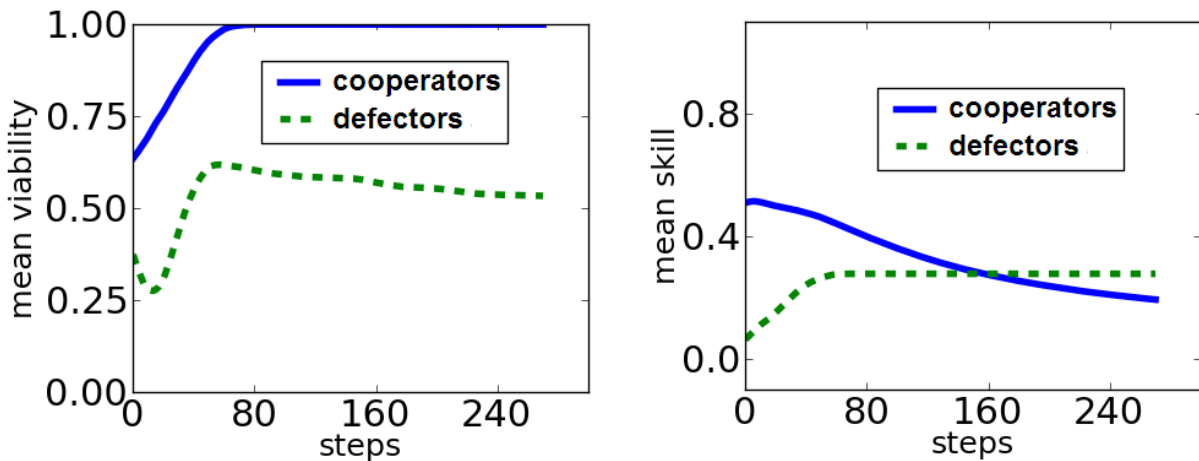
Figure 5: The left panel shows the viability dynamics for both cooperators and defectors in a society of affective learner agents over 300 steps. The right panel shows the skill dynamics for both cooperative and sociopathic affective learner agents.

The outcome of this experiment replicate the results by Epstein in a qualitative way: "[...] cooperation can emerge in a population of tag-less agents playing zero-memory fixed strategies of cooperate or defect" ([6], p. 207). We can attribute this to the quick formation of a large cooperative zone. Within this zone, there is no room to move for the cooperators, hence cooperators are

decoupled from the environment, since they cannot execute their action plans. We hypothesize that clusters of cooperators inhibit the self-reinforcing behavior of affective agents. This way, the social dynamics in the DPD paradigm are only caused by the spatial constraints, and not by the behavioral choices and interaction among agents in the society.

Conclusion

The agent-based social simulation based on *affective learner* agents integrates behavioral autonomy, affect research, and game theory to achieve a more biologically plausible model of human group behavior. The model is an attempt to integrate the notions of affect-based decision-making [8, 12] and biological agency [4, 2] to the field of social simulation [6, 10]. The results showed that the affective learner model replicates the results of the DPD model. However, because of the spatial interaction model we were not yet able to fully show the power of the affective learner agent.

Our model operationalizes the life-mind continuity thesis by a simulation inspired by insights from artificial autonomy, to understand one specific social phenomenon. We showed that the complex model is able to replicate the results of the more simple DPD model. We propose the affective learner model is stronger in that it is able to connect to areas like cognitive science, sociology or ethology, while the DPD paradigm convines itself to game-theoretic approaches to social simulation.

As mentioned, a current weakness of the affective learner model is that it uses high-level abstractions to build the bridge from agency to sociality. Furthermore, the spatial framework blocks the behavior of affective agents. Future work should apply the affective learner model to different interaction models, such as graph-based network models, which enable the agents to act according to their self-reinforcing behavior. Furthermore, future iterations of the model should aim to implement agents which exhibit the ability to self-produce their boundaries. Finally, we aim to apply the affective learner model to simulate other sociological phenomena and compare it with other existing social simulation models.

References

[1] R. Axelrod and W.D. Hamilton. The evolution of cooperation. Science, 211(4489):1390, 1981.

[2] X.E. Barandiaran, E. Di Paolo, and M. Rohde. Defining agency: individuality, normativity, asymmetry, and spatio-temporality in action. Adaptive Behavior, 17(5):367, 2009.

[3] A. Clark. Mindware: An introduction to the Philosophy of Cognitive Science. Oxford University Press Oxford, UK, 2001.

[4] E.A. Di Paolo. Autopoiesis, adaptivity, teleology, agency. Phenomenology and the Cognitive Sciences, 4(4):429–452, 2005.

[5] M.D. Egbert, X.E. Barandiaran, E.A. Di Paolo, and C.V. Rao. A minimal model of metabolism-based chemotaxis. PLoS Computational Biology, 6(12):574–580, 2010.

[6] J.M. Epstein. Zones of cooperation in demographic prisoner's dilemma. Complexity, 4(2):36–48, 1998.

[7] J.M. Epstein. Generative Social Science: Studies in Agent-based Computational Modeling. Princeton University Press, 2007.

[8] N.H. Frijda. The Emotions. Cambridge University Press, Cambridge, MA, 1986.

[9] T. Froese and E.A. Di Paolo. Sociality and the life–mind continuity thesis. Phenomenology and the Cognitive Sciences, 8(4):439–463, 2009.

[10] G.N. Gilbert and K.G. Troitzsch. Simulation for the Social Scientist. Open University Press, NY, 2005.

[11] H.R. Maturana and F.J. Varela. Autopoiesis and Cognition: The Realization of the Living. Springer, London, 1980.

[12] J.A. Russell. Core affect and the psychological construction of emotion. Psychological Review, 110(1):145–172, 2003.

Metabolism and Behaviour in Simple Dissipative Structures

Nathaniel Virgo

nathanielvirgo@gmail.com
University of Sussex

Summary

I present an overview of my previous work on the study of simple (precarious, individuated) dissipative structures as a scientifically useful metaphor for living systems, focusing in particular on the analogues of metabolism and behaviour that can be found in such systems. I discuss an example in which a self-maintaining "control system" can be added that modulates the behaviour of a simple self-maintaining system. This control system is metabolically costly but pays for itself by generating beneficial behaviour, providing an interesting analogue to an animal's nervous system. This contributes to the workshop by showing one way in which metabolism and behaviour can be modelled in the same system. Further study of these simple systems can help us refine our intuitions about these concepts and will ultimately help us to understand them in the context of real organisms.

Introduction

One aim of this workshop is to address the question "How can we model an integrated autonomous system capable of displaying both self-production and flexible, self-regulating behaviour?" In previous work (Virgo, 2011; McGregor & Virgo, 2011) I argued that many simple dissipative structures (i.e. self-maintaining physical systems such as hurricanes) share a number of important properties with living organisms, including analogues of metabolism and adaptive behaviour. In this extended abstract I will bring the argument about behaviour into the foreground by studying a particular example.

This work fits into a general methodology of studying a class of physical systems that include living organisms but which also includes simpler structures that are easier to model and to understand. This class should be broad enough to include easily modelled examples but narrow enough not to be trivial. In this and previous work I focus on what I call precarious, individuated dissipative structures (PIDS).

This terminology is spelt out more thoroughly in (Virgo, 2011), but briefly, dissipative structures (in the sense of Prigogine, e.g. 1978) are those that arise in physical systems out of thermal equilibrium, "precarious" implies the possibility of death — a precarious dissipative structure can be irreversibly destroyed — and individuation means the separation of dissipative structures into distinct individuals, spatially differentiated from one another and from their environment, analogously to the separation of living matter into individual cells and organisms. Figure 1 illustrates the relationship between these properties and various types of physical structure, including life.
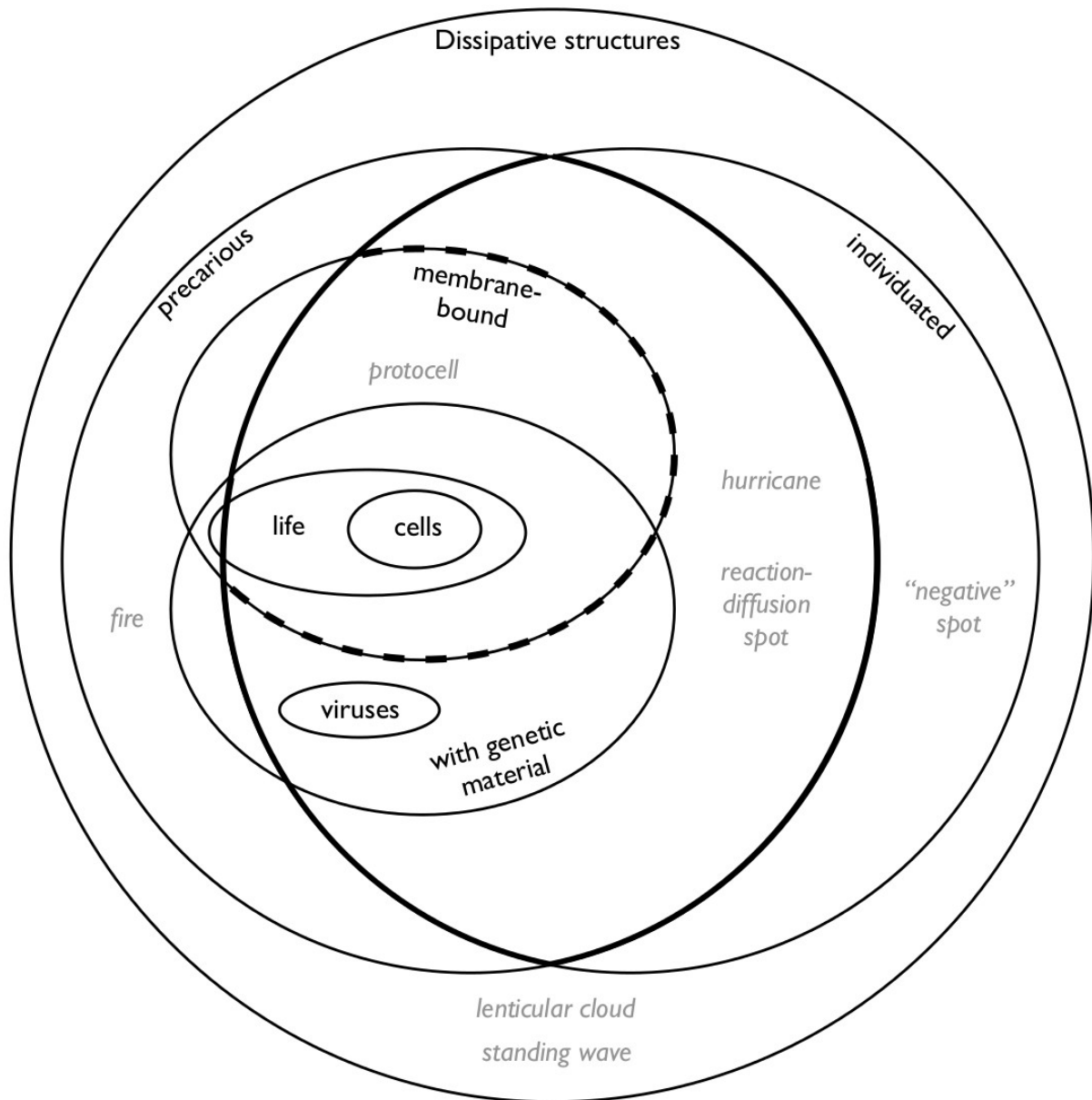
Figure 1: A Venn diagram showing the relationships between various properties of dissipative structures. Examples are shown in grey italic text. The phrase "negative spot" refers to the holes in the pattern in Figure 2e. The region corresponding to precarious individuated dissipative structures is highlighted with a thick outline; the dotted line delimits the subset of this which corresponds to the definition of autopoiesis in Maturana and Varela's later works. Life is shown extending outside of the 'individuated' region because of the existence of a few species which can lose their individuation and spread out indefinitely. However, it should be understood that the vast majority of life is individuated.

This definition of PIDS is similar in spirit to Maturana and Varela's (1980) notion of autopoiesis, although it is not compatible with later definitions of autopoiesis (e.g. in Maturana & Varela, 1987), which turn on the generation of a surrounding membrane. Colleagues and I have argued in the past that the importance of the surrounding membrane in autopoiesis has been overstated (Virgo, Egbert, & Froese, 2009).

Many previous models that combine metabolism and behaviour in the same system (e.g. Suzuki & Ikegami, 2004; Egbert & Di Paolo, 2009) also focus on the generation of a membrane. Here I will present a model using membrane-less PIDS that nevertheless exhibit adaptive behaviour, suggesting that individuation *per se* is more important than membrane generation as a defining feature of autonomous systems.

In (Virgo, 2011, Chapter 5) I applied this methodology to simple structures that arise spontaneously in a type of system known as a reaction-diffusion system. These systems model a scenario in which chemical reactions take place on a surface. As well as reacting, the reactants move across the surface by diffusion. The reactions are sustained by a continual influx of a "food" chemical into every point on the surface. A particularly simple reaction-diffusion system capable of exhibiting interesting behaviour is the Gray-Scott system, defined by the reactions $A + 2B \rightarrow 3B$ and $A \rightarrow P$. A is the food chemical and B is an autocatalyst — a molecule which reacts with the food substance to produce more of itself. B decays into an inert product P, which is assumed to leave the system.

The two-dimensional version of this system was first studied by Pearson (1993), who observed that, depending on the parameters, a range of different patters could be formed. A selection of examples of these can be seen in Figure 2; the "spots" in the last two are PIDS.
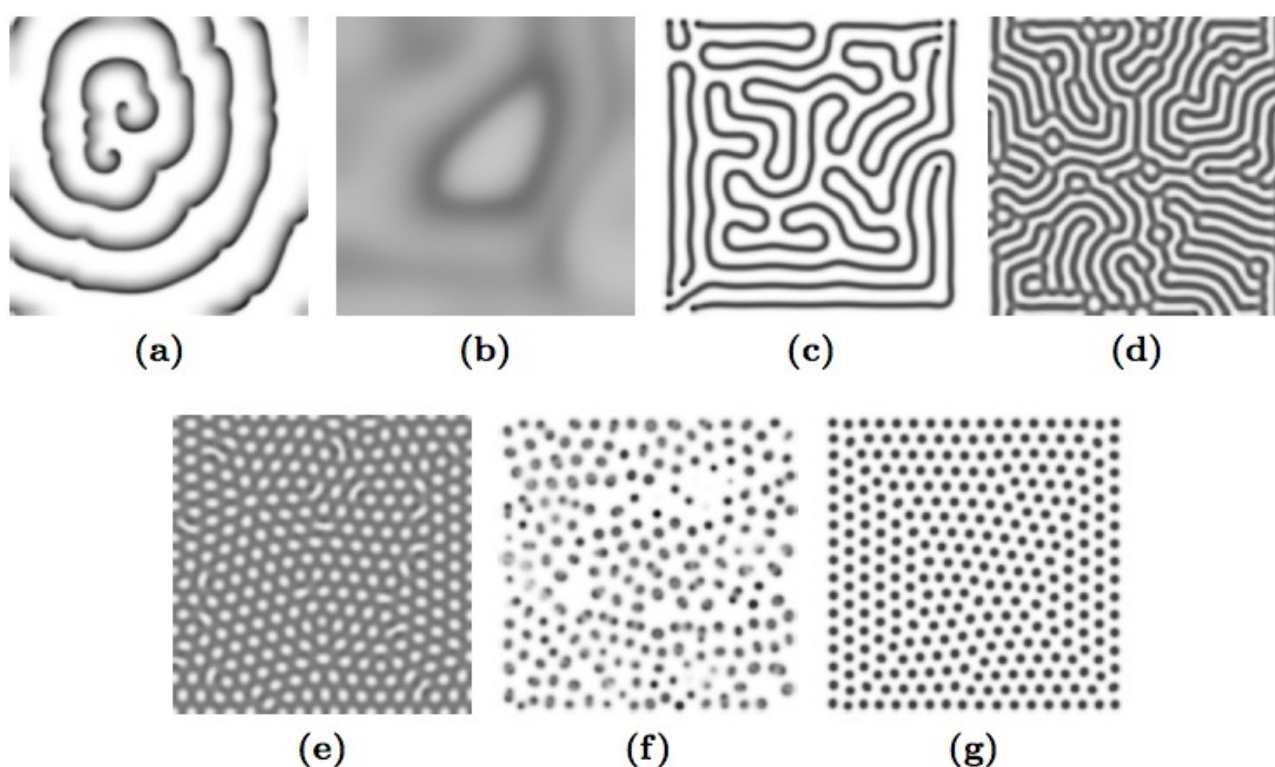
Figure 2: Examples showing the range of patterns exhibited by the 2D Gray- Scott system with various parameters. Patterns are chosen as exemplars of various phenomena; see (Pearson, 1993) for a more systematic classification. Of interest are (f) and (g), which consist of isolated "spots" of autocatalyst surrounded by regions in which little or none is present. Integration method follows (Pearson, 1993).

Metabolism

In biology, "metabolism" refers to the network of chemical reactions that take place within an organism and ultimately produce all the molecules of which it is composed. A first pass at a

definition might therefore be "a network of reactions (or processes in general) that acts to produce its own components." However, this applies to too wide a range of phenomena, since it encompasses any autocatalytic process, such as a growing crystal, fire or a nuclear chain reaction.
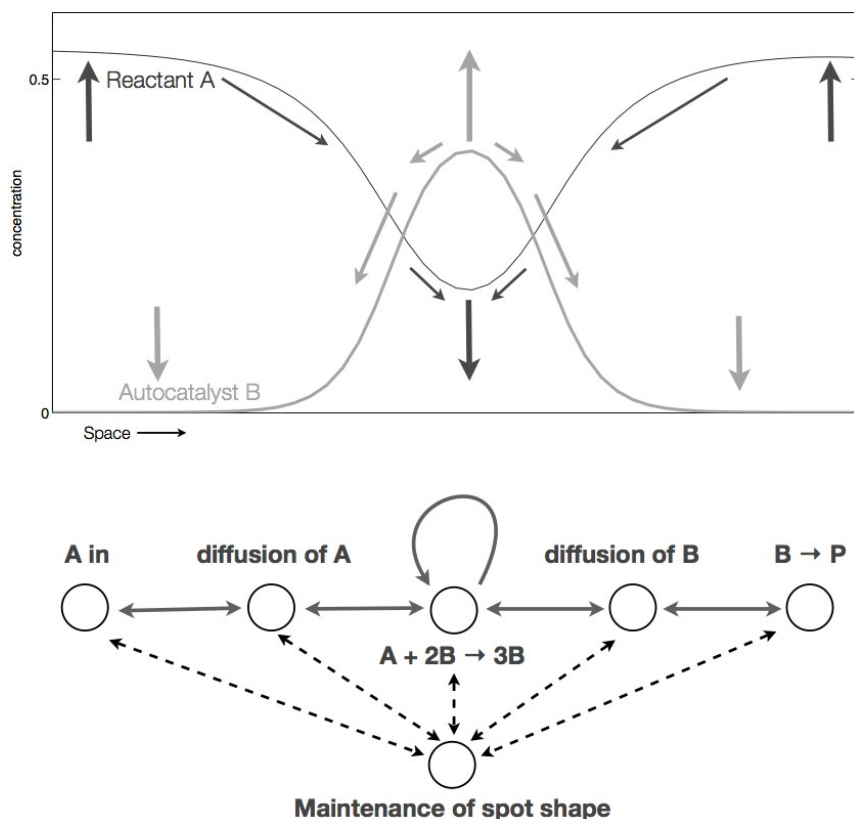
Figure 3: (top) Concentration profile across a single spot in a one-dimensional version of the Gray-Scott system. This spatial structure remains constant over time due to a balance of ongoing dynamic processes. The arrows represent the direction in which these various processes occur: vertical arrows represent the net rate at which the reactants are built up or destroyed in the system due to reactions, and angled arrows represent transport due to diffusion. (bottom) These processes can be roughly characterised as a network, with an arrow drawn between two nodes if the products of the first process enable the operation of the second. The maintenance of the spot as a whole is dependent upon all of its constituent processes, and those processes will cease if the spot ceases to be maintained.

A better definition can be given by "the network of processes that produce and maintain a single individual". Figure 3 shows how the processes that produce and maintain a reaction-diffusion spot can be considered as a network. This definition is also closely related to Maturana and Varela's (1980) definition of autopoiesis.

Behaviour

Because reaction-diffusion spots are constituted by a balance in the rates of several processes, any perturbation to these processes will cause an imbalance, which will change the configuration of the spot. Large such perturbations may destroy the spot entirely, but small enough ones will result in a return to a balanced state; the spot would not be dynamically stable if this were not the case. This return to equilibrium can, and usually does, involve a change in the spatial location of the spot.

In particular, spots tend to move up food gradients, which is part of the mechanism that keeps them individuated: if they instead moved toward regions of lower food concentration they would tend to merge together instead of remaining separate.

Stable environment | Unstable environment
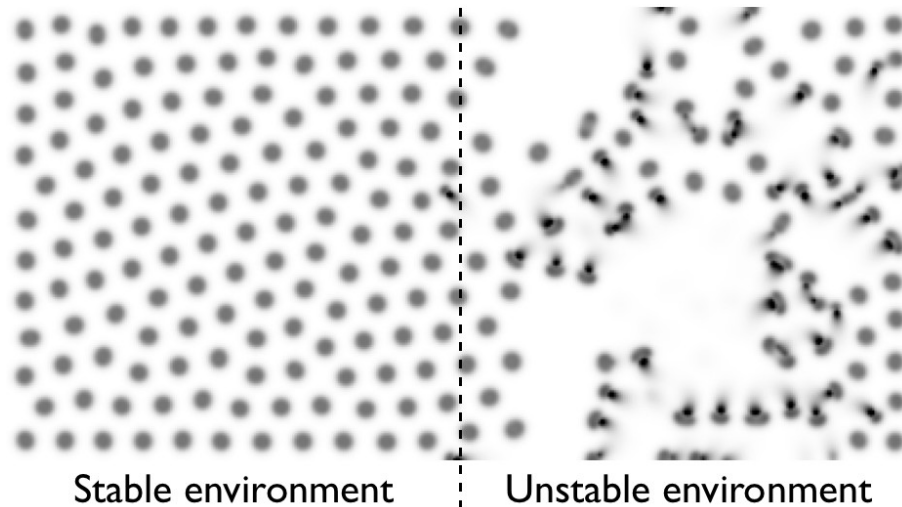
Figure 4: A snapshot of the system resulting from adding a second autocatalyst. The colours are adjusted so that the parasitic autocatalyst C, which forms into tails behind some of the spots, appears as a darker shade of grey than the primary autocatalyst B. The spots with tails move constantly away from their tails, which results in their colonising empty parts of the space more rapidly than the tail-less spots. Random areas in the right-hand side of the figure are cleared by an occasional externally-induced cataclysm. The spots with tails' ability to colonise these cleared areas more rapidly than the tail-less spots enables them to persist in this region, because their faster movement makes them better adapted to the 'niche' formed by the newly cleared areas.

A spot's behaviour can be modulated by adding a second, 'parasitic' spot that feeds on the first. A new pair of reactions, B+2C → 3C and C → P are added to the system. The dynamics of C are chosen to be slower than those of A and B in order to produce stable structures, which consist of a spot of B with a 'tail' of C. The tail eats away at one side of the spot, causing the spot to move constantly away from it. Figure 4 shows how this modulated behaviour enables spots associated with a tail of C to persist in environments where spots without tails cannot.
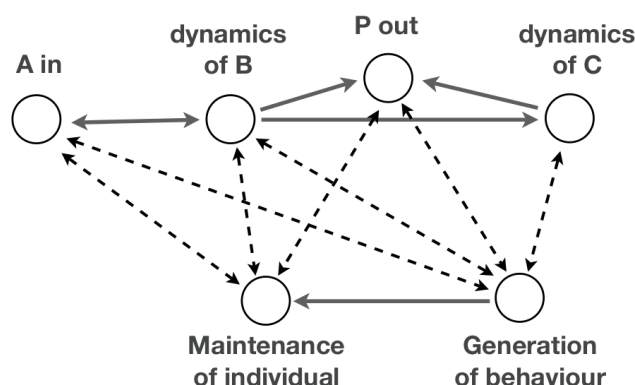
Figure 5: The network of processes associated with a spot of B together with a tail of C, again quite roughly characterised. Note that, on the 'metabolic' level, C is parasitic upon B, but, in the

appropriate environment, it nevertheless contributes to the maintenance of the individual spot+tail system through the generation of beneficial behaviour.

As Figure 5 shows, although C is metabolically parasitic, it can contribute to the maintenance of the individual by contributing to the generation of behaviour that is beneficial in a given environment. An interesting parallel can be drawn between this and the higher parts of the human brain, which is similarly metabolically costly and which contributes to the individual's metabolism through the behaviour it generates.

Conclusion

Analogues of metabolism, behaviour, and behaviour-modulating control systems can be found in simple dissipative structures. Further study along these lines can help us understand how these phenomena occur in much more complicated living systems.

References

Egbert, M., & Di Paolo, E. A. (2009). Integrating behavior and autopoiesis: An exploration in computational chemo-ethology. *Adaptive Behaviour*, 17.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht, Holland: Kluwer Academic Publishers.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. Boston, MA: Shambhala Publications.

McGregor, S., & Virgo, N. (2011). Life and its close relatives. In G. Kampis, I. Karsai, & E. Szathmáry (Eds.), *Proceedings of ecal2009, part II*, LNCS 5778 (pp. 230–237). Springer.

Pearson, J. E. (1993). Complex patterns in a simple system. *Science*, 261(5118), 189–192.

Prigogine, I. (1978). Time, structure and fluctuations. *Science*, 201(4358), 777–785.

Suzuki, K., & Ikegami, T. (2004). Self-repairing and mobility of a simple cell model. In *Artificial life IX: Proceedings of the ninth international conference on the simulation and synthesis of living systems* (pp. 421– 426). MIT Press.

Virgo, N. (2011). *Thermodynamics and the structure of living systems*. Unpublished doctoral dissertation, University of Sussex.

Virgo, N., Egbert, M., & Froese, T. (2009). The role of the spatial boundary in autopoiesis. To appear in *Proceedings of the Tenth European Conference on Artificial Life*. Springer.

Closing the Loop: From Self-Organization of Behaviour to Minimal Agency and Back

Marzena Maria Bihun (1)
Georg Martius (2,3)
J. Michael Herrmann (1,3)

(1) School of Informatics, University of Edinburgh, 10 Crichton St, Edinburgh, EH8 9AB, U.K.
(2) MPI for the Mathematics in the Sciences, Inselstr. 22, D-4182 Leipzig, Germany
(3) BFNT & MPI for Dynamics and Self-Organization, Am Fasberg 17, 37077 Gottingen, Germany

Summary

The concept of agency, as we claim here, can be reduced to a purely behavioural phenomenon. Behaviour, however, is, in this context, considered as an inseparable, but observable aspect of a self-organizing process that is guided by an optimal information flow towards the agent. Beyond a conceptual formulation, we are using this approach as a computational principle in the control of robotic agents. In the technical application the self-organising process can be characterized as the optimisation of the information flow between the active and the passive interfaces of the agent with the environment. We present results of a psycho-physical experiments where human subjects rate different types of more or less agent-like computer programs which can be observed and controlled via a narrow bandwidth channel. We find that the self-organisational algorithm is

Extended Abstract

Agency seems to arise from the dynamic interplay of organisational (noumenic) and behavioural (phenomenic) aspects which reflect processes that are, respectively, internal and the external to the agent. The principle of self-organisation, however, allows us to understand behaviour in a way that does not require an internal counterpart in order to explain the agency of a system. Moreover, self-organising processes can be described at a level of formality that permits the realisation of autonomous behaviour in robots [1]. Self-organisation of behaviour does not require a behavioural goal, a predefined set of stimuli or actions or an external source of reward, although it can be described by information-theoretic principles such as the optimality of the information flow between a robot and its environment or the rate of change of the transinformation between the environment of the robot and an internal representation of the environment in the robot. This approach has several advantages. It avoids the infinite regress which would result if the behaviour of the agent is produced by an agent inside the agent. Nevertheless, it can be hierarchically extended [2] e.g. in order to explain levels of agency, an inner perspective or to model neurophysiological data. Building on earlier computer-based experiments, we will make use here on the possibility that our approach is accessible by experimental tests. In an earlier work [3] we have compared results obtained in simulated agents with experiments [4,5] on the detection of agency which included a model of an observer of agency: The more complex the observable behaviour appeared to be the more time the artificial observer devoted to the interaction with the partner. This approach qualified the self-organising observer (or rather: explorer) as a detector of agency in a minimalistic sense. Here we will use human ratings in order to arrive at a comparison between an intuitive notion of agency and the minimalistic agency that can be attributed to the algorithm for the self-organisation of behaviour.

Analysing behaviour is essential because through the interaction or observation one can notice the factuality of agency in a situation. Yet, agent-like behaviour can be generated with just a few simple

rules and, on this level, human agency may differ only quantitatively. Nevertheless, we could hypothesise that the complexity is not enough, while the self-organisation of the behaviour might be a relevant property. To put these things together, the notion of autonomous agent can be grasped through a form of minimal agency with self-organizing behaviour and coupled internal and external aspects.

We have conducted experiments to test this hypothesis based on an established paradigm. The participants were told that they were about to interact through a computer interface with some agent trapped in a torus-shaped tunnel. They learned that the agent can move along the tunnel and they will be allowed to do so, too. The actual interaction with that agent was designed in the following way: The computer screen showed a row of three square cells representing the visible part of the tunnel. A red ball inside one of the cells represented the agent. Participants could use keyboard arrows to 'move' left or right i.e. they could change the region on the tunnel to look at. The circular tunnel consisted of eight such cells meaning that after moving into one direction for eight steps in a row, the participant would return to the same cell. The opposed agent, from now on called a runner, could also move left or right and its movements were controlled by various algorithms. A static runner did not move at all, i.e. it stayed in one cell permanently; the one-way runner was moving only in one direction at a constant pace, and the random runner was moving at each step left or right with equal probability. On top of that, to control the agent we also used interactive human input, replayed human input which was recorded beforehand and an algorithm based on the above-mentioned homeokinetic principle. In total, there were six different runners and all participants took part in six short sessions where they were naive in respect to the function of the algorithms or the experiment. They had no specific task to complete, just to explore the space and try to interact with the agent. At the end of each session they were asked to assess the agent in terms of interaction and state if they could guess the agent's strategy. As expected, participants predominantly detected the interaction with the human-controlled agent while disliking the agents that were all the time static or was moving in one way only. The results also showed that most subjects also detected the sense of interaction with the self-organising runner. Its score was higher than the ones of a random or human replay runner. These initial findings are very promising and suggest that our approach to the self-organisation of behaviour indeed captures the essential aspects of agency and interaction with other entities. One may claim that similar results could have been achieved using a rule-based agent with a fitness function trying to capture the sense of chasing the user while keeping distance and introducing some randomness to look more realistic. But such a behaviour would not be self-emergent and thus here, in this context, it is irrelevant.

The key observation here is that both, user and agent, had no predefined goal, yet some sense of interaction emerged and it was successfully detected. The experiment presented here was an attempt to test computer-based agents in an interaction scenario together with human users, while other, more specific experiments, are scheduled to be conducted in the next weeks. In the course of these experiments, the findings reported here will become more specific with respect to variants of the self-organisation algorithm, the information-theoretic complexity and the skill level of the participants and human agents.

Thus, we argue that important aspects of agency can be identified already within the horizon of the self-organisation of behaviour. This does not exclude the presence of organisational processes, but they are not considered as essential for the agency of the maintained system nor does it exclude natural agents where the interaction of behaviour and organisation is less obvious. Th formal description extends only to artificial agent which are describable as an analytical dynamical system in a meaningful way, while the more general point addressed here, might lead to a simplification in a more general sense. The work reported here is part of a project that investigates behaviour in robots and animals and studies applications in machine learning, robot learning, development, prosthetics, and interactive art.

References

[1] R. Der, M. Herrmann, and R. Liebscher. Homeokinetic approach to autonomous learning in mobile robots. In R. Dillman, R. D. Schraft, and H. Worn, editors, Robotik 2002, number 1679 in VDI-Berichte, pages 301-306. 2002.

[2] O. Gigliotta, Pezzulo, G. and S. Nolfi. Emergence of an internal model in evolving robots subjected to sensory deprivation. In: From Animals to Animats 11. Lecture Notes in Computer Science 6226, Springer Berlin / Heidelberg, 575-586, 2010.

[3] Martius, G., Nolfi, S., and Herrmann J. M. (2008) Emergence of interaction among adaptive agents. In From Animals to Animats 10, 10th International Conference on simulation of Adaptive Behavior, SAB 2008, Osaka, Japan, Proceedings, Lecture Notes in Computer Science. Spinger.

[4] Trevarthen, C.: The self born in intersubjectivity: The psychology of an infant communicating. In: Neisser, U. (ed.) The Perceived Self. Cambridge Univerity Press, Cambridge, 121–173, 1993.

27, 195–203, 2004.

A Dynamic Soundscape

Norihiro Maruyama
Geoff Nitschke
Takashi Ikegami

maruyama, geoff, ikeg {@sacral.c.u-tokyo.ac.jp}
Department of General Systems Studies,
Graduate School of Arts and Sciences
3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

Summary

This research proposes a computational model for governing the behavior of numerous sound files (*sound-swarm*) moving about a physical space. A swarm behavioral model instituting *active* versus *inactive* rules is used to play or assign new sounds thus controlling the sound-swarm dynamics. People, represented as *prey* in the sound-swarm model, can interact with the sound-swarm (the *predator*) in its physical space. The sound-swarm is attracted to such prey, where the strength of attraction is a function of the difference between sound patterns assigned to individual sound files (particles) in the swarm. As a result of this interaction, a sound-swarm potentially splits into several sub-swarms, thus generating an acoustically immersive and diverse soundscape. The research objective is not to model natural predator-prey behaviors but rather to use sound-swarms as a means to advance our understanding of autonomy.

Dynamic Soundscapes for Understanding Autonomy

The general research objective is to use the dynamics of an artificial sound-swarm in order to address the open problem of detecting and defining autonomy. This objective is motivated by the current lack of canonical measures for autonomy and agency. One specific research goal is to relate the acoustic dynamics of sound, moving through a physical space (a sound-swarm), to the role of human perception in autonomy detection. That is, we will investigate the question of whether, as human participants interact with a sound-swarm and the sound-swarm changes its acoustic properties as a result, is it possible for participants to distinguish between autonomous versus non-autonomous swarms simply by listening to them.

Addressing this question will allow experimenters to identify autonomy in the context the perceptions of those interacting with the sound-swarm. Furthermore, the perceptual experiences of the participants could then be related to differing acoustic dynamics of various sound-swarms. This in turn will allow experimenters to investigate the dynamics of acoustic adaptation underlying autonomous sound-swarms, and determine if autonomous sound-swarms share some *core set* of dynamics. This will be an important first step towards defining a set of mechanisms (in the context of sound-swarms) that result in autonomy.

In this research, we consider a sound-swarm to be a specific instance of a sound-scape, and within the purview of soundscape ecology research[Pijanowski et al., 2011a]. Pijanowski et al. [Pijanowski et al., 2011b] define a *soundscape* as a collection of biological, geophysical and anthropogenic sounds emanating from a landscape, varying over space and time, reflecting both important ecosystem processes and human activities. Soundscapes have been studied from many perspectives, including their cultural, historical, recreational, aesthetic, and therapeutic value to human societies [Schafer, 1994], [Torigoe, 2005], [O'Connor, 2008]. Soundscapes designed as an art form have also

been used to address the problem of mitigating outdoor noise pollution [Hellstrom et al., 2008].

The impact of natural and artificial soundscapes upon humans [Schafer, 1994], [Fisher, 1998] has been widely studied. However, there has been little research on computational models that control dynamic behaviors of artificial soundscapes that move through a physical environment, interacting with and adapting to the movements of human participants. Furthermore, to the best of the authors' knowledge, this is the first study that uses the dynamics of human and sound-swarm interactions to investigate autonomy.

This research applies the *Boids* computational model [Reynolds, 1987], coupled with active versus inactive rules to control the playing and renewal of sound files (*particles*), in order to control the acoustic dynamics of a *sound-swarm* moving about a three-dimensional physical space. Each particle's movement is denoted as a vector (direction of motion).

Boid swarming behavior is controlled by three simple rules: (1) Avoid collisions with neighboring particles, (2) Move towards the center of mass of neighboring particles, (3) Move in the direction that neighboring particles are moving [Reynolds, 1987]. Additionally, we imposed the following sound control rules as part of the sound-swarm dynamics. That is, (4)Playing an associated sound when the number of neighboring particles is between parameter values b1 and b2. Otherwise, the sound is not played. (5)The sound pattern is mutated if the number of neighboring particles is between the parameter values m1 and m2.

The rule (5) in the above is implemented as the follows; when a swarm particle is mutated, a new sound file is generated by randomly picking up the sound fragment from the prepared sound sources and will be attached to the swarm particle. In this way, the entire sound files will gradually change depending on how the swarm particles behave temporally.

The motivation behind this rule is to simulate memory dynamics. An active state implies that the associated sound is "retrieved" and the in-active state implies that the sound is hold but not retrieved. Which sound will be retrieved is determined by the grouping pattern and those in-active particles will commit to the retrieval by organizing a grouping but their sound will not be retrieved. The aspect of such memory dynamics simulated by a swarm modeling is a novel approach for understanding "conscious/unconcious" memory.

Each particle is defined by a sound file encoded in a given digital audio format. A high fidelity multi-speaker (7.1) system with the sound editing software system *Max/MSP*[1] and *Spats* plug-in[2] is used to give the acoustic illusion that the sound-swarm is moving about within the confines of a physical space. To avoid abrupt particle leaps between opposite ends of the sound-swarm's physical space, we used a closed boundary space in order to define the confines that the sound-swarm operates within.

Sound-Swarm Dynamics

In this research, there is no objective function. Rather, we propose that human participants experience the dynamics of autonomous sound-swarms via listening to them. Sound-swarm and participant interaction dynamics are created using a predator-prey model. For example, given multiple participants (analogous to prey), the sound-swarm may split into multiple sub-swarms in order to follow the participants as they move about the environment. Figure 1 illustrates the formation of sub-swarms in a three-dimensional space.

---

1 http://cycling74.com/

2 http://www.fluxhome.com/products/plugins/ircamspat
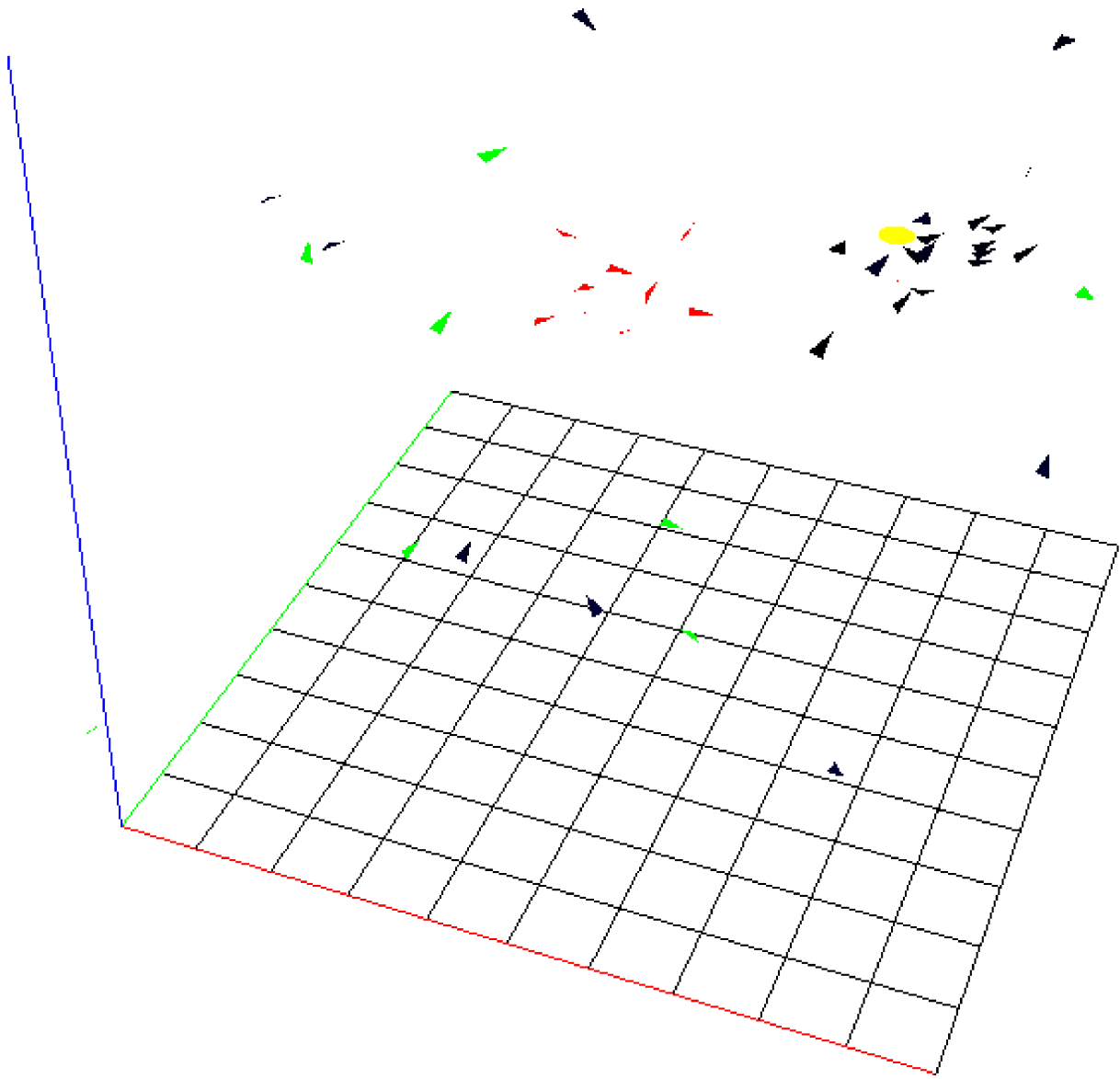
Figure1: An example of swarm dynamics with b1=3, b2=5, m1=10 and m2=30.

Red particles are in an active (sound file playing) state. Green particles are sound files that mutate. The yellow particle represents the human in the environment. Furthermore, the acoustic properties of individual particles in a swarm (or sub-swarm) will mutate whenever a participant is within a given range.

Qualitative experimental results have not yet been attained, however in the interest of defining the acoustic properties of an autonomous sound-swarm we have selected parameter values that have a *class-4* like behavior (figure 1). That is, the rules use to update the active versus inactive state of particles were inspired by class-4 type behaviors of cellular automata rules exhibited in the *Game of Life*. A unique property of this sound-swarm model is that we use active versus in-active update rules on dynamic swarm particles operating within a three-dimensional space, instead of a two-dimensional lattice. We hypothesize that such class-4 type behaviors will be an important mechanism leading to a definition of autonomy in artificial sound-swarms. Thus, we suppose that the interactions between the system and its environment creates autonomy [Ikegami, T., 2007]. That is, a system can generate and maintain its own context via temporarily coupling and decoupling

with an environmental context. Moreover, a system has its own dynamics without requiring an externally assigned task. Swarm dynamics coupled with active versus inactive state updates are an example of soundscape autonomy.

References

[Fisher, 1998] Fisher, J. (1998). What the hills are alive with: In defense of the sounds of nature. *Journal of Aesthetics and Art Criticism*, 56(1):167–179.

[Hellstrom et al., 2008] Hellstrom, B., Nilsson, M., Becker, P., and Lunden, P. (2008). Acoustic design artifacts and methods for urban soundscapes. In *Proceedings of the 15th International Congress on Sound and Vibration*, pages 423–429, Daejeon, Korea. IIAV Press.

[Ikegami, T., 2007] Ikegami, T. (2007). Simulating Active Perception and Mental Imagery with Embodied Chaotic Itinerancy. *Journal of Consciousness Studies*, Vol.14 (2007) pp.111–125.

[Ikegami et al., 2011] Ikegami, T., et al. (2011). Evolution of SoundScape with Swarm Dynamics. Submitted to SIGGRAPH Asia 2011 Emerging Technologies.

[O'Connor, 2008] O'Connor, P. (2008). The sound of silence: Valuing acoustics in heritage conservation. *Geographical Research*, 46(3):361–373.

[Pijanowski et al., 2011a] Pijanowski, B., Farina, A., Dumyahn, S., Krause, B., and Gage, S. (2011a). What is soundscape ecology? *Landscape Ecology*, In press.

[Pijanowski et al., 2011b] Pijanowski, B., Villanueva-Rivera, L., Dumyahn, S., Farina, A., Krause, B., Napoletano, B., Gage, S., and Pieretti, N. (2011b). Soundscape ecology: The science of sound in landscapes. *BioScience*, 61(3):203–216.

[Reynolds, 1987] Reynolds, C. (1987). Flocks, herds and schools: A distributed behavioral model. *Computer Graphics*, 21(4):25–36.

[Schafer, 1994] Schafer, R. (1994). *The soundscape: the tuning of the world.* Inner Traditions International Limited, Rochester, USA.

[Torigoe, 2005] Torigoe, K. (2005). Insights taken from three visited soundscapes in japan. *Soundscape*, 6(2):9–13.

Ecological autonomy: the case of a robotic model of biological cognition

Alberto Montebelli

alberto.montebelli@gmail.com

Extended Abstract

In this paper, we briefly introduce a relatively recent technology (oxygen-diffusion cathode microbial fuel cells, hereafter ODC-MFC) that has proven capable of endowing the physical prototype of a bio-mechatronic symbiotic robot with `energy autonomy'. A mathematical model of realistic ODC-MFC energy dynamics allows us to readily explore, in simulation, the cognitive/behavioral consequences for a minimalist hybrid robotic model of biological cognition provided with an artificial metabolism. The limitation of the available energy promotes the emergence of complex behavior, dynamically regulated by the essential variables of the system through a self-organized action selection mechanism that provides the system with `motivational autonomy'. In addition, the living bacterial component of the system introduces a further constraint to the viable set of behaviors. The survival of this `robotic species' is bound by the selection of behaviors that promote and maintain an ecological balance between the robotic community and its environment, over and beyond the space-time horizon of its single components. This constitutes a novel form of `ecological autonomy'.

1. The concept of autonomy, despite its importance in current robotics, often endures vague definitions. Definitions that, restricted to the focus of this paper, would approximately appear in the form: "emancipation from exogenous control of behavior". A more exacting definition of autonomy lies within the concept of autopoiesis and operational closure, as initially outlined by Maturana and Varela (Varela, 1997; Di Paolo, 2009).

2. Imagine a robot that collects water and food from its environment. These two resources are essential to the robot in order to harvest the energy it needs for sensing and action. In other words, they are crucial to its energy autonomy (McFarland, 2008). Potential food comes in the rather generic form of processed or unprocessed biodegradable mass (e.g. sucrose, rotten fruit, wastewater). The maintenance of a high level of hydration and chemical energy in its digester (both subject to temporal decay), allows for a higher rate of electric energy generation (power) that can be supplied to its actuators. On the contrary, a deficit in the water and food intake entails the incapacity to perform further action (death).

3. The above is more than a thought experiment. Oxygen-diffusion cathode microbial fuel cells (hereafter ODC-MFCs) extract energy from biomass and water. In their anodic compartment, a population of bacteria oxidizes the molecules of the biomass and liberates electrons. The migration of H+ ions through a semipermeable membrane closes the electric loop at the cathode, where they combine with oxygen. Despite its serious limitations in terms of currently available power and capacity of actuation, an actual ODC-MFC powered prototype robot of the kind outlined above does already exist, i.e. EcobotIII (see Ieropoulos et al., 2010). Theoretical and experimental results suggest the feasibility and performance boost due to MFC miniaturization. Over the last few years, the Ecobot family scaled from 8 to 48 onboard ODC-MFCs. We can expect a future escalation in this trend, that might theoretically significantly relax the current tight energy constraints.

4. In this bio-mechatronic symbiont (Melhuish et al., 2006; Froese and Ziemke, 2009), the bacterial population that colonizes the anodic compartment of the ODC-MFC introduces in the system a living component, in the form of a self-sustained bacterial colony, thus characterized by operational closure. This enriches the hybrid robotic system with biological causal powers (Di Paolo, 2003; Ziemke, 2008). Although obviously non-autopoietic, this particular robotic instance appears more relevant to the study of biological autonomy than more traditional robots. The stability of the robot's metabolism-based energy supply is crucially mediated by its behavior (Froese and Ziemke, 2009), thus revealing a form of 'extended life' (Di Paolo, 2009). As we will see in the rest of this brief introduction, the living component projects short to long term behavioral consequences onto the system, as well as the potential for a novel and special kind of robot autonomy.

5. A computationally inexpensive mathematical model of ODC-MFCs (Montebelli et al., 2011), allows researchers to anticipate the expected evolution of the ODC-MFC technology, and explore in a less constrained manner the inherent properties of this energy autonomous bio-mechatronic hybrid. For example, we can readily study a large number of onboard ODC-MFCs, contemporarily powering our robot in simulation, whilst maintaining realistic energy dynamics. A neurocontroller can be implemented by an artificial neural network, whose weights are identified via evolutionary algorithms. A fitness function might be a rather generic requirement such as "live as long as you can" (e.g. see Montebelli et al., 2010). Which kind of observations can we expect from the self-organization of this setup? The robot will be under selective pressure for the development of effective strategies for the tracing and exploitation of the environmental resources (food and water). In the case of evolvable robot morphologies (not yet tested), also the body could be adapted to these goals. More sophisticated sets of sensors might implement simple chemical analysis, promoting the development of preferences that would be related to the intrinsic energy content of the available food. The variables that are essential to the viability of the system (food and water levels) work as its control parameters (Kelso, 1995), enacting a dynamic action selection mechanism that elicits a number of qualitatively different behavioral attractors (e.g. see Montebelli et al., 2008; Montebelli et al., 2010). This implements a form of motivational autonomy (McFarland, 2008). Furthermore, more complex controllers might develop energetically efficient behaviors, for example by activating, prior to the direct engagement in action, less energetically demanding sets of actuators for an initial screening of the scene (as in Lowe et al., 2010) or abstract planning and thought. In addition, bursts of high power activation, leading to cycles of energy accumulation and mobilization, or lower power continuous actuation would be two feasible behavioral options (e.g. see Montebelli et al., 2010). Finally, in future experiments, the robotic agent might learn how to spatially organize areas dedicated to its foraging and excretions (the latter being provided with fertilizing properties), and temporally rotate them to achieve a better harvest. All of these neural, morphological and behavioral consequences directly derive from the ODC-MFC induced energy constraints, emerging within the possible forms of robot-environment coupling.

6. We can consider the ODC-MFC system powering the robot as a simple first approximation of a robotic artificial metabolism. Interpreting the robotic model as a metaphor of a biological agent, the low-frequency metabolic internal dynamics (the essential variables that control the agent's motivational autonomy) associate the contingent and high-frequency sensorimotor flow (that characterizes the interaction of the agent with its environment) with the non-negotiable essence of adaptivity: the organismic well being. Blindness to this primary fact amounts to pursuing a myopic perspective on cognitive research, trapped in contingent and local dynamics, whilst ignoring that cognition amounts to nothing but the deployment of a sophisticated strategy for survival.

7. Now imagine to increase the number of onboard ODC-MFCs enough, so that the available energy is sufficient to virtually support any demand by the powered robot over an extended period of time. In which sense would the described system differ from robots supplied with more traditional forms of energy (e.g. batteries)? Firstly, the system would not be subject to the same thermodynamically irreversible processes that characterize more traditional power sources. In principle, the anodic bacteria constitute a rather robust and dynamically self-sustaining population. Secondly, and most importantly, an ODC-MFC powered robot will still crucially depend on the resources at hand in its environment. It would be viable as long as its behavior promotes a balanced and sustained relation with its environment. Behaviors that are disruptive of the ecological balance within the space-time horizon of this robotic species would be unreconcilable with its collective long-term viability. In other words, the viable robots would be ecologically grounded in their environment and their specific form of autonomy would be constrained by the maintenance of their ecological balance. By ecological autonomy we mean a form of energy and collective motivational autonomy that mainly depends on the demands of the agents' viable integration in their natural environment over time.

References

E. Di Paolo. Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In K. Murase and T. Asakura, editors, Dynamical Systems Approach to Embodiment and Sociality, p. 19–42. T, Adelaide, 2003.

E. Di Paolo. Extended life. Topoi, 28:19–42, 2009.

T. Froese and T. Ziemke. Enactive artificial intelligence: Investigating the systemic organization of life and mind. Artificial Intelligence, 173:466–500, 2009.

I. Ieropoulos, J. Greenman, C. Melhuish, and I. Horsfield. EcoBot-III : a robot with guts. In H. Fellermann et al. (eds.), Artificial Life XII, p. 733–740. T, 2010.

J. A. S. Kelso. Dynamic Patterns: The Self-Organization of Brain and Behavior. The MIT Press, 1995.

R. Lowe, A. Montebelli, I. Ieropoulos, J. Greenman, C. Melhuish, and T. Ziemke. Grounding Motivation in Energy Autonomy: A Study of Artificial Metabolism Constrained Robot Dynamics. In H. Fellermann et al. (eds.), Artificial Life XII, p. 725–732. T, 2010.

D. McFarland. Guilty robots, happy dogs: the question of alien minds. Oxford University Press, 2008.

C. Melhuish, I. Ieropoulos, J. Greenman, and I. Horsfield. Energetically autonomous robots: food for thought. Autonomous Robots, 21:187–198, 2006.

A. Montebelli, C. Herrera, and T. Ziemke. On Cognition as Dynamical Coupling: An Analysis of Behavioral Attractor Dynamics. Adaptive Behavior, 16(2-3):182–195, 2008.

A. Montebelli, I. Ieropoulos, R. Lowe, T. Ziemke, C. Melhuish, and J. Greenman. An Oxygen-Diffusion Cathode MFC Model for Simulation of Energy-Autonomous Robots. Submitted.

A. Montebelli, R. Lowe, I. Ieropoulos, C. Melhuish, J.Greenman, and T. Ziemke. Microbial fuel cell driven behavioral dynamics in robot simulations. In H. Fellermann et al. (eds.), Artificial Life XII, number 1960, p. 749–756. T, 2010.

F. J. Varela. Patterns of life: intertwining identity and cognition. Brain and cognition, 34(1):72–87, 1997.

T. Ziemke. On the role of emotion in biological and robotic autonomy. BioSystems, 91:401–408, 2008.

A quantum theory of biological autonomy at the macro-level

Tom Froese
Takashi Ikegami

t.froese@gmail.com
Ikegami Laboratory,
Department of General Systems Studies
University of Tokyo, Japan

Abstract

There is a peculiar irony at the core of biology that in order to fully understand an organism we have to kill it and dissect it. While the existence of an organism depends on self-determination and holistic integration, our knowledge of it depends on external control and decomposition. There is a deep tension between acquiring knowledge of another autonomous being and the persistence of that being's very autonomy. But if complete knowledge of the living is in the end impossible, because the *explanans* ultimately requires the destruction of the *explanandum*, can there ever be an adequate science of life? We propose that a theory of biological autonomy that is modeled on the mathematical framework of quantum physics might be a workable alternative.

Introduction

Although marginal to mainstream biology, over the years there have been a number of prominent proposals to overcome the aporia that *biology*, the study of life, is in fact premised on *thanatology*, the study of death. The authors of these proposals typically try to compel us to bracket the power structures, which are intrinsic to the control/knowledge construct, by adopting an enabling discourse centered on autonomy/understanding. We can find some aspects of this kind of alternative approach to biology in the writings of, for example, von Uexküll, Jonas, Maturana, Varela, Rosen, Goodwin, and others. All of these authors share the conviction that some form of organizational closure is essential to the phenomenon of life. It is this closure of the organism's defining organization that partly constitutes the possibility of its existence as an autonomous individual.

Several formalisms and mathematical models have been proposed to capture the notion of organizational closure with varying degrees of success. Nevertheless, they do not allow for the essential ambiguity that must necessarily arise out of the tension between our external knowledge and the other's autonomous being. In order to illustrate this shortcoming we can recall one of the earliest attempts to formalize organizational closure, namely Ashby's concept of 'ultrastability' (see Froese and Stewart 2010 for a more extensive discussion).

Beyond Ashby: A self-breaking system

Ashby realized that a defining feature of all organisms is their ability to adapt to changing and novel circumstances, and he linked this adaptivity with changes to their internal organization. In brief, he posited boundaries of viability that, if exceeded, would lead to a new organization due to parametric changes in the organism's old organization. This model of adaptation is based on a mechanism of random change, which Ashby called 'breaking'. He demonstrated that when a system is perturbed to exceed its boundaries of viability, it continues breaking until a new stable organization had been found. The system then ceases to do anything until it is again sufficiently disrupted to trigger the break mechanism.

At first sight Ashby's formulation of ultrastability may also seem to be a useful model of autonomy,

given that the ultrastable system is capable of changing its organization. However, as Ashby was keen on emphasizing, it is also possible to view the living system and its environment as one whole system in which there are no 'breaks' and the state of all variables is fully known at every time step. But this presence of complete knowledge about the system violates the prospect that the model can adequately capture a living being's autonomy. In the words of the phenomenological tradition we could say that if the phenomenon can be fully captured by this kind of system, then, although it is closed in one sense, it is still lacking its own 'transcendence' or intrinsic otherness. It does not have the ability to escape from the scope of our awareness to constitute an alternative center of subjectivity. Similar worries can be raised about other theories of autonomy, which also grant that the whole system can be fully known by us in principle.

In order to get an intuitive idea about what type of organization would be more suitable for describing a living being in systemic terms, let us consider the possibility of a *self-breaking* system. Like in Ashby's terminology, we use the term 'break' to specify a change in the system's organization, but in this case the 'break' is applied by the system to itself. On this view, a living system can be conceived as a self-sustaining process of changing organization, which continues changing the way it is organized until it dies. Here we therefore have inverted Ashby's concern with the conditions for a *static stability* by an insistence on the need for a *dynamic instability*, and we have displaced the source of this ongoing instability to become an essential part of the organism.

This notion of a self-breaking organization ensures that a living being cannot be fully known in principle. In order to precisely know the *state* of a living system's organization we need to measure it a precise point in time, but this means that we cannot accurately measure its *change* in organization since that requires a measurement over an extended period of time. Conversely, if we want to know how its organization changes over time, we cannot simultaneously know its precise state at any one point in time. The upshot of this uncertainty principle is that the being of the organism always escapes our knowledge of the organism to some extent. We submit that this is a necessary condition of autonomy.

And yet this uncertainty principle of autonomy is just the first step: it only gives us a 'shallow' uncertainty based on the fact that a living system's organization is continually in flux. This uncertainty has to do with a fundamental issue of observer-relative measurement, and it does not say much about the intrinsic organization of the system. If we want to conceptualize organisms as necessarily self-breaking systems then we need a deeper sense of uncertainty in order to do justice to their autonomous existence. But this leaves us with a methodological and conceptual problem. How can we replace the certainty of death with the uncertainty of life as the ultimate foundation of biology?

Deep uncertainty: Toward a quantum theory of autonomy

Fortunately, science and mathematics have already provided us with some of the tools we will need, although they have been developed in a different context: quantum physics. Indeed, quantum physics is also faced by the problem that its phenomena are necessarily unknowable to their full extent. And yet this inherent uncertainty has not prevented it from becoming one of the most precise and practically useful scientific disciplines. The basic trick is to formalize the very uncertainty of the phenomenon as a superposition of its various possibilities, and then to conceive of a measurement as temporarily revealing a slice of these possibilities according to a distribution of probabilities.

We suggest that we should construct a quantum theory of autonomy on the basis of a similar kind of formalism. A living being is then no longer fully described by having any one organization, but rather by a superposition of various possible organizations that have different likelihoods of becoming materially manifest at each point in time. This is not as odd as it may sound. It is well known in the field of artificial life that adding an element of chaos to a modeled organism gives its

behavior a more lifelike appearance. The quantum theory of autonomy has the same effect, except that here the unpredictability is intrinsic to the phenomenon, rather than merely being due to the addition of an external source of noise.

We do not pretend to have a fully worked out formalism of autonomy derived from the mathematics employed in quantum physics, but at least on a general level of description we highlight the following features:

Observer-relativism: most of cognitive science proceeds under the highly questionable assumption that knowledge of other agents is independent from the observer (i.e. it is a representation, not an interaction). But this theory supports a different view: what you perceive depends on what you do! It follows that agency detection, for instance in primatology and infant studies, can no longer be conceived in a decontextualized manner.

Non-commutability: most theories of autonomy appear to assume that the processes underlying closure are symmetrical in time, such that the arrow of time can be reversed without effect (Di Paolo 2005). But this theory, again, supports a different view: there is an inherent directionality to the sequence of events! And this makes intuitive sense when describing the behavior of agents. Consider, for instance, that laughing and then being angry has a different meaning than being angry and than laughing.

Quantization: Ever since Bergson and Husserl's analysis of our experience of temporality we know that time has a temporal duration; it is not simply a list of infinitesimally small steps. But, as Di Paolo (2005) has argued, organizational closure as such does not entail duration. And yet on this view there may be a certain minimal unit of spatiotemporal existence.

Of course, this is just the very beginning of a quantum theory of autonomy, and much remains to be done to practically work out the details. And we do not deny that there may be other formalisms that are more suitable for the job. However, we strongly suggest that autonomy necessarily entails uncertainty, both for the autonomous agent itself and for its external observer. The main message of our proposal is therefore this: no formal theory of biological autonomy can properly do justice to the phenomenon, if it does not allow for the fact that life involves an essential uncertainty at its very core.

References

Di Paolo, E.A. (2005), "Autopoiesis, adaptivity, teleology, agency", *Phenomenology and the Cognitive Sciences*, **4**(4), pp. 429-452

Froese, T. & Stewart, J. (2010), "Life after Ashby: Ultrastability and the autopoietic foundations of biological individuality", *Cybernetics & Human Knowing*, **17**(4), pp. 7-50

The idea of closure in autonomous systems

Maël Montévil[1]
Matteo Mossio[2]

[1] montevil@di.ens.fr, LIENS, ENS and Paris V, FDV doctoral school
[2] matteo.mossio@ehu.es, Department of Logic and Philosophy of Science, University of the Basque Country, Spain

Abstract

In this paper, we provide a general characterization of closure as a distinctive causal regime of autonomous systems, which cannot be reduced to physico-chemical causation without loosing relevant information about the system's organization.

Our argument consists in three steps. First, we put forward an account of how different levels of causation can be realized within a biological autonomous systems, by making a conceptual and formal distinction, based on the idea of symmetry, between processes and constraints exerted on these processes. Second, we develop the notion of dependence among constraints and, third, we claim that closure is realized as a mutual dependence among a set of entities having the status of constraints within the system. The paper might then make a relevant contribution to the elaboration of a conceptual and formal theory of closure, able to overcome some of the weaknesses of previous accounts.

Introduction

The notion of closure is usually considered as one of the central determinations of autonomy. According to Varela, (operational or organizational) closure designates an organization of processes such that "(1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist" (Varela, 1979, p. 55). According to the "Closure Thesis", closure is a necessary requirement for autonomy, to the extent that "every autonomous system is operationally closed" (Varela, 1979, p. 58).

During the last three decades, substantial accounts of closure, in addition to Varela's one, have been proposed in the field of Theoretical Biology (Pattee, 1982; Rosen, 2005; Kauffman, 2002), and scientific investigations on closure has been developed in various directions (Chandler & van de Vijver, 2000). In spite of the increasing amount of scientific work on closure, however, existing conceptual and formal characterizations have a crucial weakness, since it is still unclear whether or not closure constitutes a distinctive causal regime of autonomous systems, and whether or not it might be reduced to physico-chemical causation without loosing crucial information about the system's organization.

In this paper, we will advocate the view according to which closure can be characterized as a distinctive and irreducible causal regime of biological systems, interpreted as a typical class of autonomous systems. Our strategy will consist in three steps. First, we will put forward an account of how different levels of causation can be realized by biological systems, by making a conceptual and formal distinction, based on the idea of symmetry, between *processes* and *constraints* exerted on these processes. Second, we will develop the notion of dependence among constraints and, third, we will claim that closure is realized as a *mutual* dependence among a set of entities having the status of constraints within the system.

The outcome of the paper will be a account of closure in which it is possible to distinguish, within

an autonomous system, between a causal level of open thermodynamic transformations and a causal level of closed dependence among constraints. In particular, on the basis of the explicit formulation of the requirements that entities must fulfill in order to have the status of constraint, the account makes a relevant contribution to connect in a more straightforward way the theoretical expression with empirical aspects of relevant biological phenomena.

Constraints and processes

The claim according to which organizational closure is a distinctive causal regime of autonomous (biological) systems requires an account of the specificity of closure with respect to other kinds of causal regimes in the natural world. In particular, given the dissipative nature of biological systems, which are traversed by a flow of matter, energy and entropy through their boundaries, the relations between (thermodynamic) openness and (organizational) closure should be clarified in theoretical and formal terms.

In this paper, we account for the distinction between levels of causation in terms of the distinction between processes and constraints (exerted on these processes). Processes refer to the whole set of physico-chemical changes (including reactions) occurring in biological systems, which involve the alteration, consumption, and/or production of relevant entities. Constraints, in turn, refer to entities which, while acting upon these processes, can be said to remain unaffected by them, at least under certain conditions or from a certain point of view.

We propose to ground the theoretical and formal distinction between processes and constraints on the concept of symmetry, which in general terms indicates a transformation that doesn't change the relevant aspects of an object (Weyl, 1983; Goodman & Wallach, 2009; Bailly & Longo, 2011). In particular, given a process $A \rightarrow B$ (getting B from A), C is a constraint on $A \rightarrow B$ if and only if the following three conditions are fulfilled:

I/ The situations $A \rightarrow B$ and $A_C \rightarrow B_C$ (the former under the influence of C) are not symmetric by permutation at a time scale τ.

II/ A temporal symmetry is associated to C with respect to the process $A_C \rightarrow B_C$, at the time scale τ

III/ The space of possible evolution of $A_C \rightarrow B_C$ is smaller or equal to the space of possible evolution of $A \rightarrow B$, each space being described at the relevant scale.

Conditions I-II-III can be met either directly or after analytic approximation. The situation which fulfills conditions I-II-III will be expressed as $A \rightarrow^C B$ (τ) or, in an expanded form:
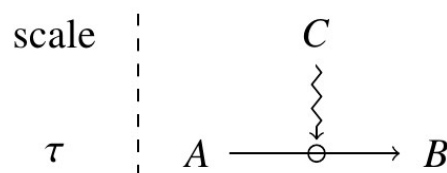
Let's discuss the three conditions by referring to a concrete example, i.e. the vascular system taken as a constraint on the flow of oxygen in an organism.

First, there is an asymmetry associated with the flow of oxygen when considered under the influence of the vascular system ($A_C \rightarrow B_C$) or not ($A \rightarrow B$) since, for instance, $A_C \rightarrow B_C$ occurs as a transport (canalized) to each cell whereas $A \rightarrow B$ has a diffusive form. As a consequence, the situation fits condition I, which means that the vascular system exerts a causal role on the flow of oxygen.

Second, a temporal symmetry is associated to the vascular system C with respect to the transformation $A_C \rightarrow B_C$ since, among other things, the topology of the vascular system remains unaltered during the blood flow. Hence, the situation fits conditions II, which means that C is conserved throughout the process.

Third, the flow of oxygen could reach each cell at an adequate rate even in the form $A \rightarrow B$, i.e. in the absence of the vascular system, from the point of view of statistical mechanics. Hence, the vascular system doesn't extend the space of possible evolutions of the process $A \rightarrow B$ which means that the situation fulfills condition III. In other terms, the vascular system is not required, at least in principle, for oxygen to reach the cells at an adequate rate — the probability of the unconstrained situation to occur being, however, very low.

Since the three conditions are fulfilled, the vascular system can be taken as a constraint on the flow of oxygen, and constitutes, we claim, a distinct level of causation.
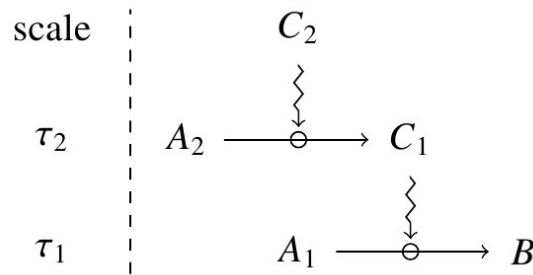
It is crucial to emphasize that each condition is met *only at the relevant time scales*. Let's take condition II, in our example. Although the topology of the vascular system doesn't change for small time scales, it does change on greater time scales when, for instance, neovascularization occurs and changes topological aspects of the system. The same holds for enzymes, which can be shown to fit conditions I-II-III, and then play the role of constraints. Enzymes are not consumed during the process they catalyze, while they decay and can randomly disintegrate at larger scales. At a very small time scale, enzymes are not conserved too, since they bind with the substrate, lose or gain hydrogen or electrons, etc.. Yet, there is a relevant scale in which, when the reactive process is achieved, the enzyme is restored, leading to symmetry.

Note, in addition, that the time scale τ at which conditions I and II must be fulfilled is the same. In turn, for condition III, the relevant scales at which the space of possible evolutions of $A \rightarrow B$ and $A_C \rightarrow B_C$ can be described can be very different from τ.

Dependence

Entities acting as constraints within an autonomous system may have a relation of "dependence" among them.

As we mentioned, each constraint C is associated to a time symmetry at the scale of the process that it constrains (τ_1 below), but not necessarily at other scales (τ_2). At the same time, C can itself be the product of a process which, in turn, can be constrained by another constraint. This situation leads to the following diagram of minimal causal dependence among constraints:

$$
\begin{array}{c|ccc}
\text{scale} & & C_2 & \\
& & \downarrow & \\
\tau_2 & A_2 & \longrightarrow\!\!\!\circ\!\!\!\longrightarrow & C_1 \\
& & \downarrow & \\
\tau_1 & & A_1 \longrightarrow\!\!\!\circ\!\!\!\longrightarrow & B
\end{array}
$$

More generally, let's consider a constrained process $A_1 \rightarrow^{C_1} B \,(\tau_1)$. Because of condition II, there is a time symmetry at scale τ_1 associated to C_1. At the same time, C_1 is the product of another constrained process $A_2 \rightarrow^{C_2} C_1 \,(\tau_2)$, at a different time scale. At scale τ_2, C_2 plays the role of constraint, whereas C_1 does not, being the product of the process $A_2 \rightarrow^{C_2} C_1$. In this situation, we claim that there is a *dependence* of constraint C_1 on constraint C_2.

More precisely, we define a relation of dependence among constraints a causal regime in which, given two time scales τ_1 and τ_2 considered jointly:

- It exists an object C_2 which is a constraint at scale τ_2 on a process producing C_1
- C_1 is a constraint at scale τ_1

A couple of important remarks: First, dependence among constraints is logically distinct from

34

dependence among processes. In fact, at scales other than τ_2, where C_2 does not play the role of constraint, there might be no other causal connection among the processes involved.
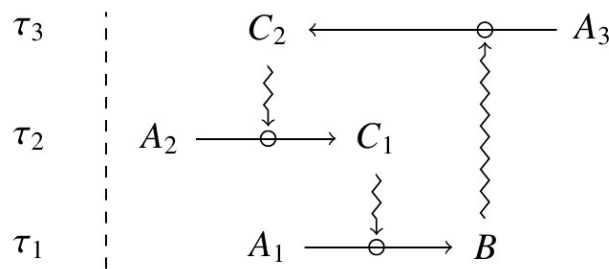
Second, because of condition III, $A_2 \rightarrow C_1$ does not require C_2 for occurring, at least at the very large scale of its possible evolutions, which implies that dependence can be observed only at the specific scale τ_2.

It is noteworthy that "repair mechanism" can be analyzed as dependence among constraints (Wang, Michelitsch, Wunderlin, & Mahadeva, 2009). On the one hand, repairing implies the existence of a subsystem (C_1) which is only marginally altered while the main process occurs (i.e. its alteration is negligible at the relevant scale, τ_1), although it is altered on the long run (τ_2). On the other hand, the maintenance of the system's organization requires, at time scale τ_2, the existence of a second subsystem (C_2) in charge of maintaining C_1 through the adequate canalization of a process $A_2 \rightarrow^{C2} C_1$.

Closure

The third step of the paper will consists in defining closure as a specific case of dependence among constraints such that, given a set of constraints C, each constraint of C depends on, and contributes to the maintenance of, (at least) another constraint of C.

Let's consider a simple example of a system realizing a closure among constraints:

$$
\begin{array}{c|ccc}
\tau_3 & & C_2 \longleftarrow & \!\!\!\!\!\!\!\circ\;\; A_3 \\
& & \wr & \wr \\
\tau_2 & A_2 \longrightarrow \!\!\!\circ\!\!\longrightarrow & C_1 & \wr \\
& & \wr & \wr \\
\tau_1 & & A_1 \longrightarrow \!\!\!\circ\!\!\longrightarrow & B
\end{array}
$$

As for whatever relation of dependence, two scales must be considered for each constraint (C) subject to closure: a scale ($\tau_s(C)$) at which C is associated to a time symmetry, and a scale ($\tau_d(C)$) at which it is produced and/or maintained. As shown in the diagram, closure requires that at least one constraint for which $\tau_s(C) - \tau_d(C) > 0$ and another constraint for which $\tau_s(C) - \tau_d(C) < 0$.

As we will discuss, however, these requirements do not necessarily lead to complex relations between scales. More precisely, we will distinguish between two possible situations. First, it might be the case that, when the constrained processes are regular enough, the apparently complex "scale structure" of the system is in fact understandable at the upper scale alone (scale separation). Second, the multi-scale dynamic might be sustained as such, which would lead to an irreducibility of the multi-scale description (West, Bologna, & Grigolini, 2003).

The outcome of the paper is a characterization of organizational closure based on an explicit formal distinction between two levels of causation (constraints and processes) whose relations, in turn, rely on the identification of symmetries at the relevant (temporal) scales. Accordingly, the paper could then provide a relevant contribution to the elaboration of a conceptual and formal theory of closure as a distinctive causal regime of autonomous organization, able to overcome some of the weaknesses of previous accounts. Yet, although the specific formulation developed in this paper is supposed to capture some of the *distinctive* features of organizational closure (with respect to whatever other causal regime), it cannot be taken, *per se*, as a definition of the closure at work in biological autonomous systems. The complexity of biological organization would require more accurate accounts, capturing the additional properties of closure as realized by full-fledged biological systems[4].

---

4 To mention an important example, biological organization (in contrast with other dissipative physical phenomena)

References

Bailly, F., & Longo, G. (2011). *Mathematics and the natural sciences; the physical singularity of life.* Imperial College Press, World Scientific.

Chandler, J., & van de Vijver, G. (2000). *Closure: emergent organizations and their dynamics.* New York Academy of Sciences.

Goodman, R., & Wallach, N. (2009). *Symmetry, representations, and invariants.* Springer Verlag.

Kauffman, S. (2002). *Investigations.* Oxford University Press, USA.

Pattee, H. (1982). *Cell psychology: an evolutionary approach to the symbol-matter problem.* Cognition and Brain Theory, 5(4), 325–341.

Rosen, R. (2005). *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life.* Columbia University Press.

Varela, F. (1979). *Principles of biological autonomy.* North Holland New York.

Wang, J., Michelitsch, T., Wunderlin, A., & Mahadeva, R. (2009). *Aging as a consequence of misrepair–anovel theory of aging.* Arxiv preprint arXiv:0904.0575.

West, B., Bologna, M., & Grigolini, P. (2003). *Physics of fractal operators.* Springer Verlag.

Weyl, H. (1983). *Symmetry.* Princeton University Press.

---

require that some constraints subject to closure determine the boundaries of the system, typically in the form of a semi-permeable membrane.