Can evolutionary robotics generate simulation models of autopoiesis?

Tom Froese and Ezequiel A. Di Paolo

CSRP 598

December 2008

ISSN 1350-3162

Cognitive Science
Research Papers

Can evolutionary robotics generate simulation models of autopoiesis?

Tom Froese[1] and Ezequiel A. Di Paolo

Centre for Computational Neuroscience and Robotics (CCNR)
Centre for Research in Cognitive Science (COGS)
Ì½»¨¾«Ñ¡, Brighton BN1 9QH, UK

[1]E-mail: t.froese@gmail.com

Abstract

There are some signs that a resurgence of interest in modeling constitutive autonomy is underway. This paper contributes to this recent development by exploring the possibility of using evolutionary robotics, traditionally only used as a generative mechanism for the study of embodied-embedded cognitive systems, to generate simulation models of constitutively autonomous systems. Such systems, which are autonomous in the sense that they self-constitute an identity under precarious conditions, have so far been elusive. The challenges and opportunities involved in such an endeavor are explicated in terms of a concrete model. While we conclude that this model fails to fully satisfy all the organizational criteria that are required for constitutive autonomy, it nevertheless serves to illustrate that evolutionary robotics at least has the potential to become a valuable tool for generating such models.

Keywords: artificial life, autopoiesis, evolutionary robotics, autonomy.

Note: This manuscript was originally targeted at the audience of this year's *Artificial Life XI* conference, which was held in August in Winchester, UK. Though it ended up being rejected both as a paper and as an abstract, it might still be of interest to others, and is therefore made available here with only minor alterations.

Prologue

It is the year 1991, Paris, and in their introduction to the first *European Conference on Artificial Life*, the organizers Francisco Varela and Paul Bourgine issue a stern warning to the field: "We reiterate that in order to avoid falling into the trap of a mere fashionable buzz word, or a fascination with technological wizardry without direction, it is important not to lose sight of the deep issues that animate this resurgence of research. Our view, as we stated at the outset, is that AL finds its *élan* because it (re-)discovers the central role of the basic abilities of living systems as the key to any form of knowledge […] our stance is that autonomy is the emblematic quality which needs to be unfolded into clear and practical concepts" (Bourgine & Varela 1992).

The conference series is thus inaugurated with a clear message: the defining goal of artificial life research is to gain a better scientific understanding of the kind of autonomy that is characteristic of living systems. Bourgine and Varela (1992) conclude their introduction with the rather optimistic outlook that "theoretical, conceptual and engineering progress is quite possible in notions that until recently were dismissed as merely metaphorical" and that "the practice of autonomous systems is not any longer a matter of mere vague speculation in contrast to a well developed theory of control systems". Today, 17 years later, we again find ourselves in Europe in the context of the *Artificial Life* conference series which, for the first time in its 20 year history, is held outside the USA. We would like to take advantage of this occasion to ask: was Bourgine and Varela's warning heeded? Was their optimistic outlook warranted? Are we any closer today to a well developed theory of autonomous systems? How can we know?

1. Introduction

In this paper we want to address the question of whether we can we take advantage of the progress that has already been made in synthesizing and analyzing the dynamics of embodied-embedded cognition in order to advance our understanding of constitutive autonomy. In particular, we will analyze the suitability of evolutionary robotics as a method to generate models of constitutively autonomous agents. We first address some theoretical and methodological issues, and then discuss their implications in terms of a concrete simulation model. While this model fails to satisfy all the necessary criteria for constitutive autonomy, it nevertheless points toward the possibility that our proposed re-conceptualization of evolutionary robotics might provide a way of overcoming many perceived shortcomings of using this method.

1.1 Two conceptions of autonomy

How much progress has been made toward a better understanding of autonomous systems in the artificial life community? The problem is that the answer to this question depends to a large extent on what we mean by 'autonomy'. However, a recent literature review of artificial life research has confirmed what many people in the field may already suspect, namely that there is no widely accepted definition of autonomy (Froese, et al. 2007). In order to address some of the confusion which this ambiguity entails, a conceptual

distinction between *behavioral* and *constitutive autonomy* was advocated in that paper. The former is intended to capture the fact that the notion of 'autonomous robotics' has essentially become a synonym for any methodology aimed at synthesizing artificial 'agents' with situated behavior for practical engineering or scientific purposes. In this context the label 'autonomy' is generally used to indicate that the behavior of the system is independent of the experimenter in some relevant sense. The notion of constitutive autonomy, on the other hand, was introduced to capture a much more specific use of the term, namely when it is used to refer to the ability of certain systems to self-constitute an identity under precarious conditions. Research in this latter context is generally more focused on understanding autonomy in relation to biological systems.

Which of these two uses of the concept 'autonomy' did Bourgine and Varela have in mind? If we interpret "Towards a practice of autonomous systems", which is the title of their introductory paper and the slogan of the European conference series as a whole, as referring to the practice of synthesizing artificial systems for the study of situated and embodied cognition, then there has clearly been progress toward the establishment of just such a research program (e.g. Harvey, et al. 2005; Beer 2003). But are the systems that are produced in this manner actually models of *autonomous systems* in the sense which Bourgine and Varela originally intended?

That this is not the case is clear from the way in which they introduce the notion of autonomy in relation to actual living creatures, which leads them to claim that "autonomy in this context refers to their basic and fundamental capacity to *be*, to assert their existence and to bring forth a world that is significant and pertinent without being pre-digested in advance" (Bourgine and Varela 1992). Moreover, as an example of the conceptual unfolding of 'autonomy' they refer to the "Closure Thesis", which states that every autonomous system is operationally closed. Varela (1979, p. 55) provides us with an explicit description of this view:

> An autonomous system can be defined in operational terms as a system with an organization that is characterized by 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)

The paradigmatic example of such constitutive autonomy is found in the chemical domain in the form of the metabolic self-production of the living cell, an organizational property which has come to be known as *autopoiesis* (Maturana and Varela 1980). How much progress has been mode toward a practice of such *constitutively* autonomous systems?

1.2 Reappraising the progress in artificial life

Unfortunately, Bourgine and Varela's (1992) original vision for the artificial life community has been diffused over the years. Nevertheless, there exists a small but

dedicated group of researchers in the field who specifically engage with the challenge of modeling biological autonomous organizations.

Since the paradigmatic example of constitutive autonomy is the living cell it should come as no surprise that many attempts of producing a model of the minimal biological organization have focused on simulations of primitive cells in simplified chemical domains (e.g. Ono and Ikegami 2000; Fernando 2005). Interestingly, work in this area has already begun many years before the proper inception of the field by Langton (1989) when Varela, Maturana and Uribe (1974) developed a cellular automata model of autopoiesis. This approach has given rise to a tradition of computational autopoiesis (McMullin 2004), and it has been shown that even the investigation of very simple oscillatory cellular automata structures can be useful in explicating some of the theoretical and conceptual issues of autopoiesis (e.g. Beer 2004). In addition, the original model by Varela and colleagues has recently been expanded to include self-movement (Ikegami and Suzuki 2008), as well as being extended to three dimensions (Bourgine and Stewart 2004). Moreover, a more realistic study of the origins of minimal cells is beginning to be possible with the development of more plausible models of artificial chemistry (e.g. Ruiz-Mirazo and Mavelli 2008) and attempts to synthesize autopoietic systems with actual chemistry (Luisi 2003).

More recently, the situation has started to look even more hopeful, as evidenced for example by two special journal issues devoted to the topic of autonomy (Barandiaran and Ruiz-Mirazo 2008; Di Paolo 2004). These special issues are especially valuable contributions because they demonstrate that the methodological toolbox for synthesizing and understanding autonomy is being expanded in new directions.

Nevertheless, despite these important efforts it is still the case that there has been relatively little progress on the problem of constitutive autonomy, especially when compared to the impressive advances that have been made in synthesizing and understanding the behavioral dynamics of artificial cognitive systems (e.g. Harvey, et al. 2005; Beer 2003). Of course, that there has been a shift of focus away from Varela and Bourgine's original vision for the field does not necessarily entail that researchers have become consumed by a "fascination with technological wizardry without direction" (Bourgine and Varela 1992). On the contrary, the focus on cognition in the artificial life community has clearly led to valuable insights into the dynamics of adaptive behavior, and has also improved our theoretical understanding of dynamical systems in general (e.g. Beer 2003; 1997). These are not only important advances in their own right, but the mathematical tools which are being developed to analyze the complex dynamics of artificial cognitive systems are bound to be helpful for the study of the dynamics of constitutive autonomy as well.

However, there is also a problem: it appears that the kinds of methods which the field uses to synthesize artificial cognitive systems are unsuitable to generate constitutively autonomous systems (Froese, et al. 2007). Indeed, it is standard practice in evolutionary robotics, which has for many researchers become the generative mechanism of choice to produce cognitive systems of interest, to abstract away from the constitutive autonomy of

biological systems. While this abstraction is undesirable from a scientific point of view because cognition and constitutive autonomy are deeply intertwined in all living systems, it is nevertheless necessitated by the fact that "a complete account of this situation would require a theory of biological organization, and the theoretical situation here is even less well developed than it is for adaptive behavior" (Beer 1997).

It thus appears that Varela and Bourgine might have been slightly too optimistic when they characterized artificial life as a field ready to be committed to the explication of a well developed theory of biological autonomy. Indeed, despite some important advances, today such a theory of biological organization is still in need of significant further development and concretization, in particular through the formulation and analysis of theoretical models (Beer 2004).

2. Methodological issues

Given that our understanding of how to synthesize and analyze simulation models of complex dynamics has advanced faster for models of minimal cognition compared to constitutive autonomy, it is natural to ask whether we can use insights from the former to move the latter forward. This section therefore begins by outlining some challenges that can be raised against the possibility of using evolutionary robotics to generate models of constitutive autonomy. For this critique we will draw on the extensive work on biological autonomy done by the San Sebastian group led by Alvaro Moreno (e.g. Moreno, et al. 2008; Barandiaran and Moreno 2006; Moreno and Etxeberria 2005; Ruiz-Mirazo and Moreno 2004; Ruiz-Mirazo and Moreno 2000; Moreno and Ruiz-Mirazo 1999; Moreno, et al. 1997). We then introduce a novel way of conceptualizing evolutionary robotics as a more general generative mechanism and argue that this subtle shift in perspective can potentially help us to address some of the methodology's perceived shortcomings.

2.1 Problems with evolutionary robotics

One of the main projects of the San Sebastian group has been to attempt a naturalization of the concept of autonomy by developing a biological account that is well grounded in the universal laws of physics and chemistry. Starting from a detailed consideration of the special material and energetic requirements of metabolism (e.g. Moreno and Ruiz-Mirazo 1999), they introduce the notion of *basic autonomy* to denote any system which has the capacity to manage the flow of matter and energy through it so that it can, at the same time, regulate, modify, and control: (i) internal self-constructive processes and (ii) processes of exchange with the environment (Ruiz-Mirazo and Moreno 2004). This conception of 'basic autonomy' leads them to the claim that the success of attempts to create artificially minimal autonomous systems is strongly linked to efforts of creating simple metabolic systems (Ruiz-Mirazo and Moreno 2000).

Accordingly, evolutionary robotics is rejected as a viable methodology, because it does not deal with systems whose physical organization is self-modifiable, in favor of artificial synthesis of chemical systems (Ruiz-Mirazo and Moreno 2004). More precisely, it is claimed that the difficulty with the evolutionary robotics approach is its reliance on

building blocks which are constitutively inert aggregates, since the material structures which support the operational level of computer simulations are entirely passive (Moreno and Ruiz-Mirazo 1999).

Evolved artificial systems can thus never achieve (full) constructive closure because the inertness of their building blocks entails that the required external degree of design complexity must always be greater than the internal one. In natural systems this is not a problem because such systems always start with "building blocks endowed with certain interactive capacities, derived from their material structure, that is to say, with intrinsically active elements whose combinations may generate new forms of activity" (Moreno and Etxeberria 2005). Moreover, since evolutionary robotics does not make explicit the complex underlying material organization of living systems, it cannot lead to models which include the thermodynamic requirements necessary for basic autonomy (Moreno and Ruiz-Mirazo 1999). This leads them to conclude that basic autonomy cannot be realized but from a highly complex chemical organization and that, as a consequence, we should not expect that work in evolutionary robotics will generate forms of agency similar to that in living ones (Moreno and Etxeberria 2005).

One way to respond to these considerations is to point out that the notion of 'basic autonomy' is actually only concerned with one particular kind of constitutive autonomy, namely the metabolic self-construction of living systems. As such we can accept their criticism of evolutionary robotics in the sense that it is not the method of choice for synthesizing actual living systems. However, we will later argue that there is no *a priori* reason why it cannot be used as a more general generative method for the creation of *models* which make explicit the requirements of a material organization – a model, after all, should be measured by its usefulness in helping to improve the understanding of a given problem even when it fails to capture essential elements since often this very failure can be informative.

Another possible response, and the one which we will develop more concretely in this paper, is to argue that a supposed failure in terms of 'basic autonomy' does not rule out the possibility that evolutionary robotics might still be a suitable method for generating other forms of constitutive autonomy. One particularly attractive target, for example, is the constitutive autonomy found in the cognitive domain of the nervous system (Varela 1991). While it is of course the case that the cognitive abilities of living systems are deeply intertwined with their metabolic self-construction (Moreno, et al. 1997), it might also be possible to give an account of cognitive autonomy that is decoupled from such material requirements.

Barandiaran and Moreno (2006) have recently proposed such a "minimally cognitive organization program" which focuses on the organizational requirements of cognition on the basis that the nervous system is *hierarchically decoupled* from the underlying processes of metabolic self-construction. In other words, while metabolism produces and maintains the architecture of the nervous system, it nevertheless minimizes its local interference with the nervous system in such a way that we can speak of the constitution of a new dynamic domain that consists of both its internal dynamics and its embodied

sensorimotor coupling with the environment. Their attempt at specifying the requirements of constitutive autonomy in terms of a dynamic organization in the cognitive domain makes this approach especially amenable to an evolutionary robotics program of research.

2.2 Organismically-inspired robotics

There exists a line of evolutionary robotics research called "organismically-inspired robotics" (Di Paolo 2003) that advocates the necessity of using models that incorporate elements of such cognitive organization. Indeed, the introduction of homeostatic mechanisms into the evolving 'agents' has resulted in models that allow us to begin exploring the possibility of the autonomous constitution of an identity in the combined neural and behavioral dynamics of the evolved systems (e.g. Di Paolo and Iizuka 2008; Iizuka and Di Paolo 2007). In terms of our scientific understanding of biological autonomy and cognition these first examples represent a significant advance over work which solely focuses on functional aspects of these biological phenomena.

However, while the organismically-inspired approach provides an important first step because it enables us to investigate the emergence of self-maintaining dynamic cognitive structures that are comprised of neural and behavioral elements, it falls short of the full requirements for the self-constitution of a cognitive system. Barandiaran and Moreno (2006) hypothesize that "an autonomous level of normativity emerges when neural dynamics have a self-maintaining organization, i.e. when the web is homeostatic and behavior is directed towards the self-maintenance of the global stability conditions of the web (and not only of a unique dynamic structure)". The problem here is that organismically-inspired robotics still requires that the experimenter provides the evolutionary algorithm with the global identity of the system which for the purposes of the model is to count as the cognitive 'agent'.

Thus, there remains one important issue that needs to be addressed even if we do change the evolutionary robotics method so that it explicitly models material or cognitive organizational requirements. It could be argued that this method is still unsuitable for studying constitutive autonomy because the evolutionary algorithm presupposes the existence of individual 'agents' for selection and the generation of new individuals (Froese, et al. 2007). In other words, evolutionary robotics cannot be used to generate models of systems which self-constitute their own identity because what counts as an individual 'agent', i.e. what constitutes its systemic identity, is always already pre-determined by the experimenter. The main contribution of this paper is to argue that this seemingly insurmountable limitation of the method can be avoided by a relatively simple shift in perspective.

2.3 Evolutionary robotics: a re-conceptualization

How can we use evolutionary robotics to generate models of systems with constitutive autonomy if the method requires that we specify the identity of the systems that it evolves in advance? At first sight this appears to be a fundamental limitation, one which holds

independently of whether we explicitly include material and/or cognitive organizational requirements or not. What can be done?

Ironically, a solution to this dilemma becomes available as soon as we seriously accept the criticism of the San Sebastian group that the artificial systems being evolved with the use of evolutionary robotics cannot be said to be models of biological agency (e.g. Moreno and Etxeberria 2005). The solution is therefore a conceptual shift: by dropping the label 'agent' to denote what is being 'evolved' we can sidestep the fundamental problem of pre-defined identities. Instead, we conceive of what is being selected as a desired property of some kind of model *component*. Indeed, in order to further minimize any potential confusion entailed by this conceptual shift we will speak of 'optimized' rather than 'evolved' components and of 'desirable' rather than 'fit' solutions. We can thus re-conceptualize the 'evolutionary' algorithm as a more general generative mechanism, one which can be used for optimizing models of *dynamical substrates* with certain desirable properties:

This shift in perspective entails that the evaluation function can now be geared toward the optimization of a dynamical substrate with initial conditions that favor the emergence of an autonomous system which self-constitutes its own identity.

Thus, there are two important implications of this shift in perspective: (i) there is no longer any problem of the evolutionary robotics method being limited to 'agents' with pre-defined identities, and (ii) whether a particular simulation model actually includes any systems that are characterized by constitutive autonomy must be determined on a case by case basis. The second implication also presents a methodological problem. However, rather than being a fundamental limitation of the method, it presents a useful challenge in that it forces us to sharpen our conceptual requirements for identifying constitutive autonomy and encourages us to devise methods which allow us to reliably distinguish such systems.

To demonstrate the potential of this conceptual move in more concrete terms we will now analyze a recent simulation model of coordination dynamics that has been generated using the standard evolutionary robotics methodology. This will allow us to hone our intuitions about what dynamical self-constitution might or might not be.

3. The simulation model

The simulation model outlined in this section is based on recent work by Froese and Di Paolo (2008). While the original model was conceived of within the traditional evolutionary robotics framework, namely to investigate a particular aspect of the dynamics of social cognition, here we will describe it only with the terminology developed in the previous section so as to avoid any potential confusion.

In essence, Froese and Di Paolo (2008) used an evolutionary robotics method to generate a simulation model of a system, comprised of two dynamical components, that is capable

of reliably establishing and maintaining an oscillatory pattern of movement under noisy conditions. A simple schematic of the simulation model is depicted in Figure 1.

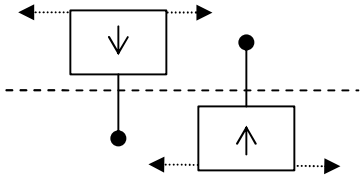

Figure 1: A schematic view of the simulation model. The two identical components are 40 units wide, only able to move in a horizontal direction, and equipped with a single on/off interface at their centre. They face each other in an unlimited continuous 1-D space. For more detailed information see text, and Froese and Di Paolo (2008).

In the evolutionary process the desirability of the model was measured according to how far away from their initial starting positions the components came into contact. This evaluation criterion optimizes the dynamics of the two components in a complex manner, namely such that their activity results in mutual localization, convergence on a target direction, and coordinated movement in that direction. Since the components are started in opposite orientation ('up' vs. 'down'), it is not possible for the evolutionary algorithm to result in the hard coding of any trivial solution (e.g. 'always move left'). In addition, this task is made even more non-trivial since 'sensory' stimulation only correlates with the overlapping of position (when the centers of the components are less than 20 units of space apart); it does not convey the direction or speed of movement of the other component. Moreover, if the components are not in direct contact with each other, the environment holds no information about their relative positions.

The basic elements of the simulation model can be described as follows: There are two components which face each other in an unlimited continuous 1-D space (i.e. one component faces 'up' and one component faces 'down'). Distance and time units of the simulation are of an arbitrary scale. Each component can only move horizontally. In terms of non-linear interaction, one on/off interface is located in the centre of each component. The interface is activated (set to 1) when the components cross each other, otherwise it is set to 0. Interface and movement noise is introduced into the simulation in order to increase the robustness of the evolved coordination pattern. In order to further increase the robustness, the initial relative displacement between the components varies between trials (range [-25, 25]).

The two components are controlled by two identical continuous-time recurrent neural networks (CTRNNs), as described by Beer (1995), each consisting of 3 fully-connected nodes with self-connections. The time evolution of the node activation is determined as follows:

$$\tau_i \dot{y}_i = -y_i + \sum_{j=1}^{N} w_{ji} z_j(y_j) + SI_i \quad z_i(x) = 1/(1+e^{-x-b_i})$$

In this equation y_i represents the activation of node i, z_i is the node output as calculated by the standard sigmoid function, t_i (range [1, 100]) is its time constant, b_i (range [-3, 3]) is a bias term, and w_{ji} (range [-8, 8]) is the strength of the connection from the node j to i. I_i represents the input to node i and S is the input gain. The total number of nodes N is set to 3; there are no hidden nodes (all nodes are affected by changes to the interface). The input is calculated by multiplying 1/0 (on/off) by an input gain parameter S (range [1, 100]), and this is applied to all nodes. There is one node, which only receives input and does not directly affect the external position, and two nodes for controlling movement; one for leftward and the other for rightward velocity. Each velocity is calculated by mapping the output onto the range [-1, 1] and then multiplying it by an output gain parameter (range [1, 50]). The overall component velocity is calculated as the difference between the left and right velocities. The time evolution of the simulation environment and each component's dynamics is calculated by using Euler integration with a time step of 0.1.

The system of components is generated by using a simple genetic algorithm (GA) which is based on the microbial GA, a steady-state GA with (rank-based) tournament selection (Harvey 2001). Until some termination criterion is reached, two solutions of the population are chosen at random, both have their desirability evaluated, and while the 'winner' of the tournament remains unchanged in the population, the 'loser' is replaced by a slightly mutated copy of the 'winner'. We define a generation as the number of tournaments required to generate a number of offspring equal to the population size. The population size is set to 40 and the algorithm terminates after 5000 generations. For a more detailed description of the evolutionary algorithm, see Froese and Di Paolo (2008).

It was possible to optimize models which are highly successful at shaping the dynamics of the components so that they come into contact as far away as possible from their initial positions (i.e. a long distance traveled together). Interestingly, the components interact in such a way that they always end up with positive *relative* displacement after their initial localization. With this arrangement the complexity of the task has been reduced considerably: while perturbation of a component's interface is ambiguous (in addition to the interference of noise, there is also no indication about the direction or speed of the other component's movement), the impact of a perturbation has now been co-organized as a 'contact on the left' indicator. This change is made possible because the dynamical systems controlling the components are not symmetric.

After the initial alignment we find that the component's coordinated movement in one direction consists of continuous oscillations induced through mutual perturbation. In other words, the velocity of each component is adjusted such that they engage in structural coupling at relatively regular intervals. It was found that this ongoing mutual perturbation is necessary for the establishment and maintenance of the coordinated pattern of movement in one common direction.

Can we account for the oscillating pattern in dynamical terms? Since the output of the 'internal' node of each component is always saturated at 1 during oscillation we can focus on the dynamics of the two 'output' nodes. If the components are not in contact

with each other ($I_i = 0$), there is a globally attracting stable equilibrium point in activation space at (-3.4, -7.5). Being in this state effectively slows down rightward velocity of component 'up'. Because of this the components eventually make contact. When $I_i = 1$ the equilibrium point is shifted to (0.3, 1.9). This effectively speeds up the rightward velocity of the component.

Interestingly, under normal conditions the dynamical system never reaches either of the two equilibrium points, because their existence is made transitory through the ongoing interaction. This is illustrated in Figure 2 in terms of the 'motor' node firing rates for component 'up' over a whole run (50 units of time). Starting from a situation of high activation of both 'motor' nodes, the system then decreases its left 'motor' firing rate in an oscillatory fashion until remains oscillating around a transitory equilibrium point.

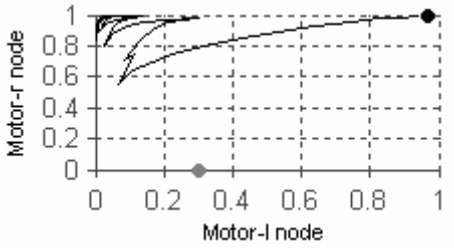

Figure 2: State trajectory of the outputs for the 2 'motor' nodes of component 'up' during mutual (two-way) interaction. The gray and black dot represent the globally attracting stable equilibrium point when sensory input $I = 0$ and $I = 1$, respectively.

But do these components act independently of each other or do they actually form a coherent system of relations? This can be tested operationally simply by recording the movement of component 'down' during a successful trial and then restarting that trial with the same initial conditions while playing back its recorded movement while component 'up' is allowed to interact as normally. It turns out that in the case of this 'playback' regime the directed coordinated movement pattern fails to be established. After some initial contact between the two components they proceed to move past each other and head into opposite directions until the end of the trial. These two situations are illustrated in Figure 3.

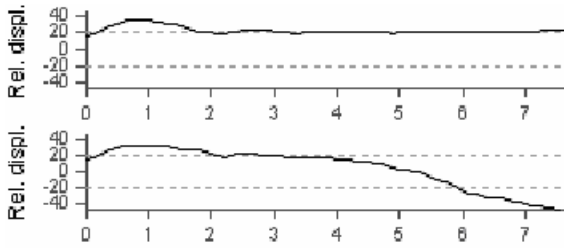

Figure 3: Change in relative displacement between the two components during the initial time steps of a trial run for two different regimes. Top: mutual (two-way) interaction. Bottom: playback (one-way) interaction.

From this we can conclude that the evolutionary process did indeed result in the generation of a system whose existence depends on the active and responsive interaction

of its two components. On their own, the components are unable to engage in oscillatory movement. Moreover, once this system has been established during the initial stages of the trial, the system displays its own (global) coherence, which constrains the (local) dynamics of the components in such a way that they both move in the same direction, in a manner that is robust to large quantities of noise.

4. Discussion

Now that we have used an evolutionary robotics methodology, re-conceptualized as a more general generative mechanism, to produce a simulation model with dynamical properties which might lead to the emergence of a constitutively autonomous system we are faced with the task of determining whether such a system can indeed be distinguished within the model. In what manner would such a system manifest itself?

Of course, since our model does not include any explicit aspects of the special material organization required for material/energetic self-construction, it clearly fails to satisfy that particular essential requirement for 'basic autonomy'. But what about the possibility of finding a system with constitutive autonomy in a domain of abstract dynamics? One important clue in this regard is that in the case of the 'basic autonomy' of metabolism "the system can achieve constructive closure because it creates high-level constraints that act on the (low-level) individual elements, harnessing their dynamics, which in turn recursively produces those control constraints" (Moreno and Etxeberria 2005). Can we find something akin to such constructive closure in the model?

4.1 A systemic analysis

If we treat the two components in our model as a systemic whole then it is clearly the case that this whole constrains the movements of the individual components. On their own they will always move in opposite directions, while in combination they move into one of the two directions together. Moreover, this constraint harnesses the dynamics of the individual components in a novel manner such that they engage in oscillatory movement. A single component will fail to coordinate with an 'inert' recording of the other component's movement (even if the conditions are the same as in the previous interaction). We are thus faced with a peculiar situation in which the oscillatory movement of the individual components brings forth the interaction process, and that interaction process enables the oscillatory movement of the individual components.

The fact that this interaction process is not only *constituted by* but also *constitutive of* the oscillatory movement of each component points to the constitutive autonomy of the interaction process. But does the organization of this system fulfil Varela's (1979, p. 55) operational definition (see quote in the Introduction)? First, we need to address the non-trivial issue of what exactly constitutes a *process* in the system. The CTRNN components are clearly not created by any activity within the model. What is created, however, is oscillatory movement. Solitary components cannot give rise to such behavior. Thus, as a first approximation we might say that the transient dynamics of each component models a process which manifests itself through an oscillatory change in position.

Second, we need to show that these two processes are related in the form of a *network*. Fortunately, this criterion is more easily fulfilled as the two processes are structurally coupled through the interface of each component. Moreover, it has been shown that the two processes *recursively depend on each other* for their generation and realization: oscillatory movement is only possible when there is (two-way) structural coupling between the two processes. It appears that we can describe the organization of the system consisting of the two oscillatory processes in such a way that it fulfils criterion (1). Does it also satisfy criterion (2)?

The problem here is that it is not quite clear what Varela means by "a unity recognizable in the space (domain) in which the processes exist". What is the unity and what exactly is the domain? Perhaps we could consider the one-dimensional environment to be the domain, but how do we distinguish a unity in that domain? One way to approach this question is by considering some of Varela's later writings on the topic of constitutive autonomy, in which he characterizes such a unity as a "selfless self" (Varela 1991): *a coherent whole that is nowhere to be found and yet can provide an occasion for the coordinated activity of ensembles of processes*. He considers this unity of coordinated activity as a point of reference for a domain of interactions in which we can distinguish the behavior of the system. While it would be possible to use the recurrent dependence between processes as the criterion to determine whether they belong or not to a dynamical 'unity', such a move would be more convincing if the model included a richer context than it currently does.

We have already stated that the ensemble of the two processes gives rise to coordinate activity and that this activity manifests itself in a particular form of behavior, namely as oscillatory movement toward a common direction. Moreover, the interaction between the two oscillatory processes provides the occasion for this common movement, but in a manner in which that 'whole' cannot be located. Its existence can only be ascertained through operational tests, for example by observing the breakdown of coordination during the 'playback' condition. Moreover, the system as whole displays a certain coherence as indicated by its robustness to large amounts of external noise (cf. Froese and Di Paolo 2008).

Due to the shift of perspective on evolutionary robotics that we have advocated we have been able to distinguish several important features associated with the organization of a system with constitutive autonomy in the evolved simulation model. Nevertheless, we suggest that the system does not fully satisfy all necessary criteria because of its impoverished domain of interactions. In particular, it is not clear that the oscillatory movement along the 1D space has a consequence for the unitary identity in any relevant sense, and the system therefore fails to satisfy Varela's criteria (2).

If we attempt to distinguish the unity in terms of it being a reference point for a behavioral domain that includes interaction with other unities, that is a proper cognitive domain, then we are at a loss because there is evidently nothing in this model with which the system as a whole can be said to interact in some way. For this reason the system also

fails to satisfy Barandiaran and Moreno's (2006) two necessary and sufficient principles of *identity* and *agency*, which characterize the existence of constitutive autonomy in the neurodynamic domain, because the latter requires that behavioral interactions result in the maintenance of the identity.

4.2 Future work

While the system we have distinguished within the simulation model falls short of being adequately describable as a simple model of constitutive autonomy, these shortcomings are not the kind of seemingly insurmountable problems that the use of evolutionary robotics is commonly believed to entail. Instead, they provide the motivations for a research program aimed at overcoming them, for example by introducing additional 'background' components into the model such that the system can interact with (and distinguish itself from) them. The next step would therefore be to think about how to change the 1D spatial environment such that it promotes the appearance of a system exhibiting global behavior which can be said to be necessary for the ongoing maintenance of the coherent systemic 'whole'. This would also enable us to investigate the relationship between constitutive autonomy and adaptivity, both of which are crucial for the sense-making abilities of living systems (cf. Di Paolo 2005).

Moreover, this approach to evolutionary robotics offers the possibility of advancing the mathematical formulation of constitutive autonomy, in particular because the dynamics of CTRNNs have already been the target of extensive study (e.g. Beer 2003; 1995). In this manner it might be possible to gain a deeper understanding of the general principles of biological organization. One possibility could be to combine this modeling approach with a descriptive formalism such as the hierarchy of dynamical systems proposed by McGregor and Fernando (2005). On the other hand, an investigation of the self-constituting system's dynamics could also be insightful. Bourgine and Stewart (2004), for example, hypothesize that the dynamics of constitutive autonomy are characterized by two 'attractors' separated by a point of bifurcation, where one 'attractor' must correspond to the disintegration of the system and the other to viable activity (see also Ono and Ikegami (2000) for a similar claim). Further work needs to be done in order to determine whether this is actually the case in the current model.

However, at first sight, an interpretation of Figure 2 suggests a radical alternative to this view. The self-maintaining dynamics result precisely from the balancing act between two (potentially many) attractors that lead both to the 'destruction' of the dynamical pattern. It is precisely because the components are *not* falling into any of the available attractors that the coherence of the system maintains itself. Perhaps a similar shift of perspective may apply to the dynamics of autonomy in general.

5. Concluding remarks

When Varela and Bourgine organized the first *European Conference on Artificial Life* in 1991 they hoped that it would push the field toward the study of the organization of biological autonomy. However, today we find that most artificial life researchers are

focused on synthesizing and understanding the behavioral dynamics of cognitive systems, while the investigation of constitutive autonomy has been largely marginalized. A large determining factor for this shift of focus is surely that autonomy, as the defining quality of all living beings, turned out to be more difficult to tackle than originally expected. However, it is time that the field of artificial life makes another concerted effort to improve our understanding of constitutive autonomy. Such an undertaking is not only desirable from the point of view of providing a strong foundation for systems biology, but is also crucial for the development and establishment of the enactive paradigm in the cognitive sciences (cf. Froese 2007).

Fortunately, it appears that a resurgence of interest in constitutive autonomy might be underway in the artificial life community. The aim of this paper was to contribute to this new focus of interest by showing that we can take advantage of the progress that has already been made in using the methodology of evolutionary robotics for synthesizing and understanding behavioral dynamics. We have argued that this can be accomplished through a simple re-conceptualization of the method as a more general generative mechanism. While the particular model that we investigated in this paper fails to fully satisfy all the organizational criteria that are required for constitutive autonomy, this study nevertheless served to illustrate that evolutionary robotics has the potential become a valuable tool for investigating this most basic biological organization.

Acknowledgements

Tom Froese wishes to thank Nathaniel Virgo and Eduardo Izquierdo for their many helpful comments and discussions.

References

Barandiaran, X. and Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior*, **14**(2): 171-185.

Barandiaran, X. and Ruiz-Mirazo, K. (2008). Introduction. Modelling autonomy: Simulating the essence of life and cognition. *BioSystems*, **91**(2): 295-304.

Beer, R. D. (1995). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, **3**(4): 471-511.

Beer, R. D. (1997). The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems*, **20**(2-4): 257-289.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, **11**(4): 209-243.

Beer, R. D. (2004). Autopoiesis and Cognition in the Game of Life. *Artificial Life*, **10**(3): 309-326.

Bourgine, P., and Stewart, J. (2004). Autopoiesis and Cognition. *Artificial Life*, **10**(3): 327-345.

Bourgine, P. and Varela, F. J. (1992). Introduction. Towards a Practice of Autonomous Systems. In Varela, F. J. and Bourgine, P., editors, *Proc. of the 1st Euro. Conf. on Artificial Life*, pp. 1-3. MIT Press, Cambridge, MA.

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

Di Paolo, E. A. (2004). Introduction. Unbinding biological autonomy: Francisco Varela's contributions to artificial life. *Artificial Life*, **10**(3): 231-233.

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

Di Paolo, E. A. and Iizuka, H. (2008). How (not) to model autonomous behavior. *BioSystems*, **91**(2): 409-423.

Fernando, C. (2005). *The Evolution of the Chemoton*. Unpublished D.Phil. Thesis, The Ì½»¨¾«Ñ¡, Brighton, UK.

Froese, T. (2007). On the role of AI in the ongoing paradigm shift within the cognitive sciences. In Lungarella, M., et al., editors, *50 Years of AI*, pp. 63-75. Springer Verlag, Berlin, Germany.

Froese, T. and Di Paolo, E. A. (2008). Stability of coordination requires mutuality of interaction in a model of embodied agents. In Asada, M., et al., editors, *From Animals to Animats 10: Proc. of the 10th Int. Conf. on the Simulation of Adaptive Behavior*, pp. 52-61, Springer Verlag, Berlin, Germany.

Froese, T., Virgo, N. and Izquierdo, E. (2007). Autonomy: a review and a reappraisal. In Almeida e Costa, F. et al., editors, *Proc. of the 9th Euro. Conf. on Artificial Life*, pp. 455-464. Springer Verlag, Berlin, Germany.

Harvey, I. (2001). Artificial Evolution: A Continuing SAGA. In Gomi, T., editor, *Proc. of the 8th Int. Symposium on Evolutionary Robotics*, pp. 94-109. Springer Verlag, Berlin, Germany.

Harvey, I., Di Paolo, E. A., Wood, R., Quinn, M. and Tuci, E. A. (2005). Evolutionary Robotics: A new scientific tool for studying cognition. *Artificial Life*, **11**(1-2): 79-98.

Iizuka, H. and Di Paolo, E. A. (2007). Toward Spinozist robotics: Exploring the minimal dynamics of behavioral preference. *Adaptive Behavior*, **15**(4): 359-376.

Ikegami, T. and Suzuki, K. (2008). From homeostatic to homeodynamic self. *BioSystems*, **91**(2): 388-400.

Langton, C. G. (1989). Artificial Life. In Langton, C. G., editor, *Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, pp. 1-47. Addison-Wesley, Redwood City, CA.

Luisi, P. L. (2003). Autopoiesis: a review and reappraisal. *Natur-wissenschaften*, **90**: 49-59.

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

McGregor, S. and Fernando, C. (2005). Levels of description: A novel approach to dynamical hierarchies. *Artificial Life*, **11**(4): 459-472.

McMullin, B. (2004). Thirty Years of Computational Autopoiesis: A Review. *Artificial Life*, **10**(3): 277-295.

Moreno, A. and Etxeberria, A. (2005). Agency in Natural and Artificial Systems. *Artificial Life*, **11**(1-2): 161-175.

Moreno, A., Etxeberria, A. and Umerez, J. (2008). The autonomy of biological individuals and artificial models. *BioSystems*, **91**(2): 309-319.

Moreno, A. and Ruiz-Mirazo, K. (1999). Metabolism and the problem of its universalization. *BioSystems*, **49**(1): 45-61.

Moreno, A., Umerez, J. and Ibañez, J. (1997). Cognition and Life: The Autonomy of Cognition. *Brain and Cognition*, **34**(1): 107-129.

Ono, N. and Ikegami, T. (2000). Self-maintenance and self-reproduction in an abstract cell model. *Journal of Theoretical Biology*, **206**: 243-253.

Ruiz-Mirazo, K. and Mavelli, F. (2008). On the way towards 'basic autonomous agents': Stochastic simulations of minimal lipid-peptide cells. *BioSystems*, **91**(2): 347-387.

Ruiz-Mirazo, K. and Moreno, A. (2000). Searching for the Roots of Autonomy: The natural and artificial paradigms revisited. *Communication and Cognition – Artificial Intelligence*, **17**(3-4): 209-228.

Ruiz-Mirazo, K. and Moreno, A. (2004). Basic Autonomy as a Fundamental Step in the Synthesis of Life. *Artificial Life*, **10**(3): 235-259.

Varela, F. J. (1979). *Principles of Biological Autonomy*. Elsevier North Holland, New York, NY.

Varela, F. J. (1991). Organism: A meshwork of selfless selves. In Tauber, A. I., editor, *Organisms and the Origins of Self*, pp. 79-107. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Varela, F.J., Maturana, H.R. and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, **5**: 187-196