Agents Recognizing Emergence

Leendert van der Torre Alfred D.M. Wan*

Vrije Universiteit Amsterdam, Department of Artificial Intelligence
http://www.cs.vu.nl/~{torre,wan}
{torre,wan}@cs.vu.nl

Abstract

Agents that recognize self organizing systems and their emergent properties have a competitive advantage. In this paper we introduce a classification of self-organizing systems based on phase transitions, macroscopic dynamic structures, evolution, self-regulation, and ultra-stability. Sometimes the agent can recognize the type of self-organizing system by observing its internal mechanisms such as feedback mechanisms or essential variables, but in general he has to interpret the system's behavior and in particular *its relation with the environment*. The agent therefore uses besides theories of self-organization also theories of cybernetics.

1 Introduction

Autonomous agents and multi agent systems are active research areas in artificial intelligence. We are interested in cognitive agents that use models of their environment, including models of other agents. In agent-based software engineering it has been recognized that the major drawback of the agent perspective is that in exceptional circumstances the locality of agents may lead to unexpected global behavior. For obvious reasons this is undesirable in software engineering (it is desirable in other areas such as for example agent-based social simulation). Jennings [7] proposes to distinguish between the agent level and a global social organizational level. The latter can be used to analyze existing systems, but it is less clear how it can be used for the development of agent systems. Our solution to the problem of unexpected emergent behavior is not to distinguish between an agent and social level, and model and analyze these two levels in separation, but to incorporate the social organizational concepts in the agent.

The main contribution of this paper is a classification of several distinct types of self-organization, originating from cybernetics and self-organizing systems. Agents may use this classification to deal with self-organization and emergent phenomena. In particular, they can discriminate between emergent phenomena and phenomena which can be explained directly from events and actions (the standard model for

*Authors are listed in alphabetical order.

the latter are causal models and probabilistic methods). Moreover, after recognizing emergence he also has to be able to take the appropriate actions. In other words, agents should be intelligent enough to deal with unexpected circumstances.

Before we start, we wish to emphasize that self-organization and emergence are neither too vague nor too complex to formalize. First, despite a sometimes confusing philosophical discussion that the second law of thermodynamics – entropy always increases – implies that self-organization is impossible [13], the theories are not too vague since they originate in physics. Second, they are not too complex, because the theories are based on simple system components (the components themselves are not described in detail, only their interaction). Moreover, as we show later, there are very simple self-organizing systems which can easily be identified, and the more complex types of self-organization can also be identified by characteristic properties.

The layout of this paper is as follows. In Section 2 we consider structure through dissipation: type 1 SOS is phase transition (Section 2.1) and type 2 SOS (Section 2.2) is macroscopic dynamic structure. Here we also briefly discuss structure through evolution. In Section 3 we discuss (pre-)biotic systems: type 3 SOS (Section 3.1) encompasses self-regulation and type 4 SOS (Section 3.2) comprises the most adaptive SOS, viz., ultrastable SOS. In Section 4 we summarize our results and we discuss how intelligent agents can benefit from them.

2 Dissipation: Imposition of external factor

Self-organization [13, 9, 12] is a process whereby the organization of a system spontaneously increases, without this increase being *controlled* by the environment or an encompassing or otherwise external system. Self organizing systems have been used to explain emergent behavior in biology such as swarm or flocking behavior, in sociology to explain social roles (e.g. roles of participants in the big brother experiment), in physics to the emergence of traffic jams, in economics to explain advantages of capitalist over communist economies, etc. In artificial intelligence self-organization and emergence have been discussed in the area of reactive systems [2, 3, 4], and recently it attracted some attention in the formalization of *social* agents [5]. The objective in this work is to understand and simulate self-organizing systems. However, they not addressed the question how an agent can recognize or deal with self-organization. It has focused on the similarities between the many forms of self-organization, instead of making a classification of its different types.

2.1 PhT-SOS: phase transitions

The first class of systems in which self-organization occurs is in systems in which a phase transition occurs, such as for example a liquid turning solid. Phase transitions are very common, but they do comprise a massive change of order in the system. If a liquid turns solid crystals may form, which presents a massive reduction in the state-space of the system, i.e., a higher order and lower entropy.

The internal mechanisms of PhT-SOS are very simple. The resulting order from a drainage of energy from the system can be explained by the properties of the elements and their mutual relations. Crystal formation in water, e.g., can be explained by the polarity of the molecules. If a phase transition has occurred, the resulting order is due to a decrease in dynamics and special properties of the elements through which the elements are arranged in a pattern.

The relation between PhT-SOS and the environment is that the environmental input decreases. When a road is blocked on the highway a traffic jam results. If a harvest fails the economy may collapse.

2.2 MDS-SOS: macroscopic dynamic structure

In MDS-SOS a macroscopic order is created, because matter or energy is fed into the system. This order is due to properties of elements comprising the medium such as the viscosity of the liquid or the polarity of the molecules or the composition of gases. Here, we describe the internal mechanisms of two classes of MDS-SOS represented in Figure 1 below: Bénard circulations and autocatalytic systems.

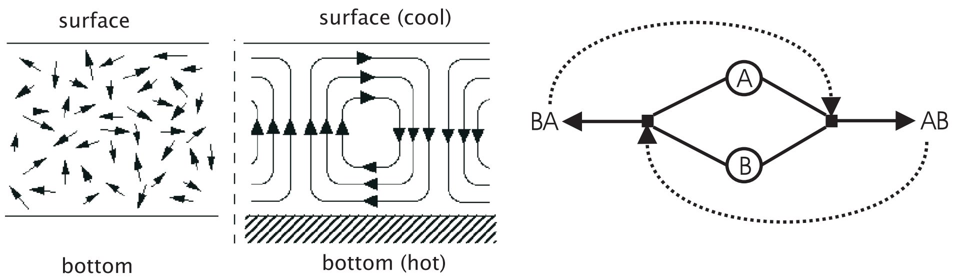

Figure 1: *On the left Bénard circulations. After heating a liquid in a container from below, hexagonal convection cells spontaneously emerge. On the right a simple autocatalytic network. In this network the reaction product AB catalyses the production of BA and vice versa.*

Bénard circulations are emergent patterns that can be observed in liquids in containers heated from the bottom, which organize themselves in convection flows of hexagonal cells (cf. [12]). “A new molecular order appears that corresponds basically to a macroscopic fluctuation stabilized by exchanges of energy with the outside world” ([12, p. 5]). Prigogine [12] calls these *dissipative structures* and show that they are at work in a great variety of systems which include self-organization in chemical reactions, regulatory processes in microbiology and cell-biology to population dynamics. The reason that dissipative structures are important, is that they comprise a separate class of mechanics, contrasting with classical dynamics. In classical dynamics, processes such as wave propagation (neglecting friction and energy loss, the process is time-independent. Under ideal conditions, a wave will continue to travel, and the phenomenon can be described by reference to a time frame which repeats itself indefinitely. Non-reversible dynamics, for instance heat conduction, concerns itself with processes that are progressive and lose their structure. As a result entropy increases. This happens when the system is isolated, but if heat and/or matter is exchanged, under certain circumstances, a new macroscopic order will occur.

Another important class of MDS-SOS are autocatalytic networks (cf. [12, 8]). Autocatalytic networks are networks of chemical reactions of which the products of the reaction catalyse their own production. When fed sufficient basic material, the reaction networks stays intact and basic material cycles through the network at a high speed. In fig. 1, a network is depicted. These networks may become very complex, because products may be produced at a large distance from where they act as a catalyst. These networks form the equivalent of a man-made factory, with simple basic materials as input and complex materials as output. Eigen [6] has shown that a result from this process, templates (RNA molecules) can emerge that contain all the information to spawn complex networks that are able to sustain themselves as well as produce the templates.

The relation between MDS-SOS and the environment is that there is an influx *and* an flux of energy out of the system through which a macroscopic dynamic structure appears. In PhT-SOS self-organization is a unidirectional process, because it incorporates dynamical patterns that 'smooth out' if they are not sustained by energy or matter influx. If the heat source sustaining the Bénard convection currents is removed, the cellular flow pattern of the water disappears, cf. fig. 1. Note that there is an external force at work which at least influences the system to exhibit a kind of order. The reason why this is not called 'control' is because the structure *itself* was not imposed on the system, but was formed as a *reaction* to the external influence or disturbance. In a more common AI vocabulary we could say that there was no explicit supervisor that imposed the order, hence, we call this unsupervised order as analogy to unsupervised learning.

3 Evolved mechanisms: (pre-)biotic systems

3.1 SR-SOS: Self-regulation

The third type of self organizing systems comprises self regulating systems, e.g., a thermostat, which can minimize the difference (the error) between actual output and desired output. For example, the human body can maintain its temperature on 37 degrees Celsius, regardless of the environment (within certain limits, of course). If the human body gets overheated, then it automatically regulates its temperature by starting to sweat (the brain has a more complex self regulating system, since it has to stay within a smaller range). This type of self-organization is related to cybernetics. The term cybernetics is due to Norbert Wiener [16], who used it to refer to the general analysis of control and communication systems in engineering as well as in biology. Cybernetic systems are a special category of self-organizing systems in which both elementary self-organizing processes, i.e., the formation of dissipative structures, as well as goal-directedness is an integral part. The self-organizational aspect of cybernetic systems is the main advantage over reactive systems, in which adaptations to new environmental conditions have to be engineered.

The internal mechanism of SR-SOS is feedback loops and therefore it is related to cybernetics. Basically, cybernetics is concerned with designing and analyzing control systems with feedback loops. From the output of a system, there is feedback to a control device that influences the environment in such a way that the output is maintained at a previously set level, or set-point. The earliest example of a cybernetic system is the governor applied in the steam machine of Maxwell [10], that regulated steam input depending on rotation of the axis it was driving (cf. [11]). In general, however, feedback does not have to run through a mechanical link. It can also, e.g., go from the central heating through the air to the thermostat that controls the power input of the central heating. Taken up in the (model of the) system are all the factors that have a direct influence on the coupling between environment and agent; all the other influences (which are usually assumed negligible) are said to be disturbances. Note that the two MDS-SOS classes depicted in fig. 1 do not contain feedback loops.

In SR-SOS, the relation between the system and the environment is much more involved than in PhT-SOS and MDS-SOS. In cybernetics the agent that acts on the environment and the environment that acts on the agent are seen as one and the same system; this is what is meant when the agent and the environment are said to be 'coupled'. There is state-determined behavior, since the state of the system at a certain point in time is determined by the previous state, though changes may occur probabilistically and there may be noise. The main feature this system has, is keeping its state within certain limits and moving it to a set-point - or a range thereof (through which it may cycle). This central feature is called *homeostasis*. In systems theory (as shown in [14]), feedback systems are widely used to produce desired output in an adaptive way.

3.2 U-SOS: Ultrastability

U-SOS comprises the availability of different types of behavior and the ability to choose from them. For example, if an animal stands in front of a fruit tree and it has a low glucose blood level, then it is able to choose behavior that causes it to pluck fruit and eat it. Certain higher animals not only have the ability to choose from behavioral sequences that are already present at birth (instincts), but are also able to *learn* new behavior.

The internal mechanisms of this adaptive behavior are made explicit in an architecture due to Ashby [1]. In this cybernetic design, there is a separation between the sets of variables with which the agent interacts with the environment, e.g. the observation of food, walking towards it and correcting the course underway, and the set with which it evaluates the adequacy of its current actions, e.g. maintaining an adequate glucose blood level. Moreover, there is a *double* feedback loop runs through the second type of variables, called *Essential Variables* (EVs). The EVs owe their name to their role in the functioning of the organism, i.e., they represent essential measures of immediate survival chances of the organism, e.g., heartbeat rate, body temperature, but also e.g., perception of a predator. If the limits of the tolerances for the Evs are crossed, the agent will switch to another kind of behavior through a *gating mechanism* because its current action pattern threatens its existence. Learning is based on trial-and-error.

In U-SOS, the relation between the system and its environment is that the system adapts its behavior, as it can be observed in animals. It can typically be characterized by a certain range of behavior of which it chooses the most appropriate behavior for the situation and its present goals. The reason this type of SOS is called *ultra*stable, is because it is stable under a range of circumstances, whereas SR-SOS is performing only *one* function, i.e., holding *one* variable stable. In par. 2.2 we have already shown that the SR-SOS evolves through a template, or genome (polynucleotides). Since the U-SOS is an extension of the SR-SOS that increases survival because it can adapt to a wider range of environments, it can be considered to be further evolved than SR-SOS.

4 Recognizing emergence

In this section we summarize the classification we introduced in the previous sections and we discuss how the agent can distinguish and *recognize* the different types of self-organizing systems and their emergent behavior. If the agent observes some phenomenon, then first he will look for a causal or probabilistic explanation, since most phenomena can be explained in that way. If he cannot find a causal explanation, then he tries to find out whether it may be an emergent property of a self-organizing system. He first distinguishes between the system and its environment, and then he discriminates the elements of the system (how to do this is outside the scope of this paper). Then the agent considers the relations between the elements of the system and the relation between the system and its environment.

		Dissipation		Evolved Mechanism	
		PhT-SOS	MDS-SOS	SR-SOS	U-SOS
type	physical	yes	yes	no	no
	biological	yes	yes	yes	yes
mechanism	group	no	yes	yes	yes
	evolution	no	no	yes	yes
	feedback	no	no	yes	yes
	essential v	no	no	no	yes
	actions	no	no	fixed	choice
behavior	environment	decrease	in- and output	complex	very complex
	set-points	no	no	yes	yes
	learning	no	no	no	yes

Table 1: Recognizing Self-Organization

In Table 1 we distinguish between four types of self-organization, two based on external factors, Phase Transition SOS or PhT-SOS and Macroscopic Dynamic Structure SOS or MDS-SOS, and two based on evolved mechanisms, Self-Regulating SOS or SR-SOS and Ultra-stability SOS or U-SOS. The systems can be distinguished if we consider how their self-organization works, i.e. by considering the relations between the elements of the system and its internal mechanisms. For example, if we observe flocking or swarming behavior it is not a PhT-SOS. If we observe evolutionary mechanisms and feedback loops the agent knows that it is not a PhT-SOS or MDS-SOS. If we moreover observe essential variables and choice of actions, then we know it is an U-SOS.

However, what should the agent do if he cannot distinguish the internal mechanisms of the system? Unfortunately, this is the case for most self-organizing systems the agent encounters. There is one basic distinction the agent can make. Since the more complex SR-SOS and U-SOS only appear in biological systems as the result of evolution, for self-organizing behavior in physical systems he can deduce this should be either PhT-SOS or MDS-SOS. But furthermore, he has to rely on an analysis of the behavior of the system.

The key concept to distinguish the different types of self-organization based on their behavior is the relation between the system and its environment. In general the rule is: *the more intimate the system is connected to its environment, the more complex type of SOS it is.* For PhT-SOS there is a sudden decrease in input, for MDS-SOS there is input *and* output, for SR-SOS there are set-points and for U-SOS there are set-points combined with learning. For example, an agent may observe a system who approaches food in his neighborhood, and who evades dangers like fire or predators; such complex interaction hints at a U-SOS. The fact that the relation between the system and its environment turns out to be the crucial discriminating factor explains the important role of cybernetics in our analysis.

5 Conclusion

In this paper we have presented a classification of self-organizing systems. We distinguished self-organization based on phase transitions, macroscopic dynamic systems, self-regulation, and ultra-stability. They can be distinguished on internal relations and mechanisms, and on the relation between the system and its environment. Our goal is to develop agents that can discriminate between emergent phenomena from self-organization and phenomena that can be explained causally. Our next step is to define different ways to anticipate or achieve emergent behavior, typically by influencing the relation between the system and its environment, and the simulation of recognition and anticipation in a multi agent system. We expect that simulations will indicate a need for further classification.

In further research we also consider whether we can use our classification for self-organization in multi agent systems, such as the emergence of social norms. The agents of a self-organizing multi agent system must be modeled as relatively simple entities, because otherwise the theories of self-organization cannot be used. For example, consider an agent who is part of a multi agent system and who has agent profiles of the other (social) agents based on cognitive concepts, such that he can compete or cooperate more efficiently and effectively, influence their behavior etc. If he models the multi agent system (of which he himself is a part!) as a self-organizing system he cannot use these agent profiles, but he has to use simpler abstractions and focus on interactions instead of social concepts. Such an agent thus models the other agents as cognitive and as reactive agents. Note that such a system is very complex since the system influences the behavior of its elements and vice versa (see e.g. [5]).

Thus, our work is a first step towards the integration of autonomous agents (AA) and multi agent systems (MAS). Our goal is a model that has four first class citizens: agents, interactions, organizations and the environment. The first two are focussed on in autonomous agents and multi agent systems, the third is focussed on in self-organization, and the fourth is focussed on in cybernetics. Moreover, our project is a first step towards the unification of artificial intelligence (cognitive and reactive agent theory), cybernetics and self organizing systems. As far as we know, these three areas have not been combined before. This unification is discussed in more detail in [15].

References

[1] W. Ross Ashby. *Design for a Brain; The Origin of Adaptive Behaviour*. Chapman & Hall, London, second edition, 1960. First published in 1952.

[2] Rodney A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics ans Automation*, RA-2:14–23, 1986.

[3] Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

[4] Rodney A. Brooks. From earwigs to humans. *Robotics and Autonomous Systems*, 20:291–304, 1997.

[5] Cristiano Castelfranchi. Modelling social action for ai agents. *Artificial Intelligence*, 103:157–182, 1998.

[6] Manfred Eigen and P. Schuster. *The Hypercycle: A Principle of Natural Self-organization*. Springer, Berlin, 1979.

[7] Nick Jennings. On agent-based software engineering. *Artificial Intelligence*, 117:277–296, 2000.

[8] Stuart A. Kauffman. *At Home in the Universe*. Oxford University Press, Oxford, 1995.

[9] Wolfgan Krohn, Günter Küppers, and Helga Nowotny. Selforganization - the convergence of ideas: An introduction. In Wolfgan Krohn, Günter Küppers, and Helga Nowotny, editors, *Solforganization: Portrait of a Scientific Revolution*, pages 1–10. Kluwer, Dordrecht, The Netherlands, 1990.

[10] Clerk Maxwell. On governors. In *Proceedins of the Royal Society*, volume 16, pages 270–283, 1867-1868.

[11] Warren S. McCulloch. Machines that think and want. In Warren S. McCulloch, editor, *Embodiments of Mind*, pages 307–316. The MIT Press, Cambridge, MA, 1965.

[12] Gregoire Nicolis and Ilya Prigogine. *Self-organization in nonequilibrium systems: From Dissipative Structures to Order through Fluctuations*. Wiley, New York, NY, 1977.

[13] H. von Foerster. On self-organizing systems and their environments. pages 31–50. Pergamon, New York, 1960.

[14] Alfred D. M. Wan, Peter J. Braspenning, and Gerard A. W. Vreeswijk. Limits to ground control in autonomous spacecraft. *Telematics & Informatics*, 12(3/4):247–259, 1996.

[15] Alfred D.M. Wan and Leendert van der Torre. Extending self-organization from system dynamics to behavior, forthcoming.

[16] Norbert Wiener. *Cybernetics, or Control and Communication in the Animal and the Machine*. Wiley, New York, 1948.