Emergence in multi-agent systems: cognitive hierarchy, detection, and complexity reductions

Dessalles, Jean Louis[1] and Phan, Denis[2]

[1] ENST, Paris, France jld@enst.fr
[2] CREM, University of Rennes I, France denis.phan@univ-rennes1.fr

1 Introduction

In a pioneering book on ”artificial society” and multi-agent simulations in social sciences, Gilbert and Conte (1995) put the emphasis on ”emergence” as a key concept of such approach: ”Emergence is one of the most interesting issues to have been addressed by computer scientists over the past few years and has also been a matter of concern in a number of other disciplines, from biology to political science” (op.cit. p.8). More recently, Agent based Computational Economics (ACE) put the emphasis on the question of emergence, following for instance Tesfatsion (2002a) or Axtell, Epstein, Young (2001). The present paper provides a formal definition of emergence, operative in multi-agent framework designed by Agent Oriented Programming, and which makes sense from both a cognitive and an economics point of view. Starting with a discussion of the polysemous concept of emergence, the first part of this paper is dedicated to clarifying the question by focussing on the problem of modelling cognitive agents in artificial societies. The key questions are introduced by way of a paradigmatic example. The second part of this paper is dedicated to introducing and discussing operative definitions and related implications. In order to illustrate our formal definition of emergence, we discusses the ACE population game model of Axtell et al. (2001) and builds a multi-level-model based on the formal framework introduced in this paper.

2 From emergentism to emergent behaviour in ACE model: some clarifications.

In this section, we first discuss different definitions of emergence, and the related background. In order to focus on the problem of modelling cognitive agents in artificial society, we next considering a paradigmatic example, and briefly discuss Schelling’s model of spatial segregation (Schelling 1969, 1971, 1978), which is a pioneering study of an emerging social phenomenon in social science.

2.1 Emergence: one word, several meanings

The notion of emergence has several meanings. In the vernacular language, emergence denotes both a gradual beginning or coming forth, or a sudden uprising or appearance; to emerge also means to become visible, in example, emergence denotes the act of rising out of a fluid. This latter sense is close to its Latin roots, where *emergere* is the opposite of *mergere*: to be submerged. In the following, we relate the ”act of rising out” to the arising of some phenomenon in a process, and note the fact that to become visible presupposes some observer. In other words, the common sense of emergence is linked to the meaning of a process that produces some phenomenon that might be detected by an observer. In the field of science, emergence has been used by Newton in optics. By the 19th century the word ”emergent” is introduced into the fields of biology and philosophy. In the latter, Emergentism has a long history, from Mill's chapter: ”Of the Composition of Causes” in *System of Logic* (1843) to the contemporary debates about the philosophy of mind, known as ”the mind - body problem” (see among others: McLaughlin, 1992, 1997, Van de Vijver, 1997, Emmeche, et al. 1997, for a synthesis). Philosophical emergentism deals with questions of both reductionism and holism. Lewes (1874) for instance places emergence at the interface between levels of organisation. For descriptive emergentism, the properties of the ”whole” cannot be defined by the properties of the parts, and results in part from some irreducible macro causal power.

In this debate around the definition of emergence, some authors have proposed to distinguish between different kinds of emergence, as for example ”nominal”, ”weak” and ”strong” emergence for Bedau (1997,2002), or ”weak” ”ontological”, and ”strong” emergence for Gillet (2002a-b). Both authors refer to debates about reductionism as well as about the so-called mind-body problem, discussing in particular the notion of Supervenience, introduced by Davidson (1970, 1980) and discussed by Kim (1992, 1993, 1995, 1999) from the point of view of emergence. As ”weak” emergence deals with upward causation and reductionism, Gillet and Bedeau relate ”*strong emergence*” to the question of ”*downward causation*” (Kim, 1992, Bedau, 2002) or ”macro-determinism”, widely advocated by Sperry (1969, 1986, 1991, among others) to deal with the mind-brain interactions, and by Campbell (1974) to deal with hierarchically organized biological systems. According to strict downward causation, the behaviour of the parts (down) is determined by the behaviour of the whole (up). For instance, parts of the system may be restrained by some act in conformity with rules given at the system level. Causation would come ”downward” in conformity with a holist principle rather than upward, according to a reductionist principle.

In this paper, we do not address theses questions directly, as we limit ourselves to discussing social behaviours in artificial societies; but the opposition between *downward* versus *upward* causation proves to be a central one in the field of social sciences. According to Granovetter (1985), the sociologist's approach would be ”over socialized” (downward) while the economist's approach would be ”under socialized” (upward/methodological individualism). Currently, both approaches have been sophisticated and are often mixed. The present paper is an attempt to integrate them in one single framework, in which the 'whole' is a collective of agents (upward causation / methodological individualism), but the agents are in are to some extent constrained by the whole (downward causation), by way of the "*social dimension*" of their belief as well as their perception of social phenomena. For the purpose of this paper, we rely on the distinction, proposed by Muller (2002) in the field of multi-agent systems, between "weak" and "strong" emergence. The latter refers to a situation in which agents are able to witness the collective emergent phenomena in which they are involved, which opens the road for both upward and downward causation.

In Agent based Computational Economics, "emergence" is strongly related to the Santa Fe approach to complexity (SFAC). In accordance with descriptive emergentism, SFAC calls "emergence" the arising at the macro level of some patterns, structures and properties of a complex adaptive system that are not contained in the property of its parts. But conversely, emergence can often be explained by upward mechanisms. Interactions between parts of a dynamic system are the source of both complex dynamics and emergence. An interesting part of the emergence process concerns the forming of some collective "order" (coherent structures or patterns at the macro level) as a result of agents' interactions within the system's dynamics, in the presence of a specific attractor. For the observer, this collective order makes sense by itself and opens up a radically new global interpretation, because it does not initially make sense as an attribute of the basic entities.

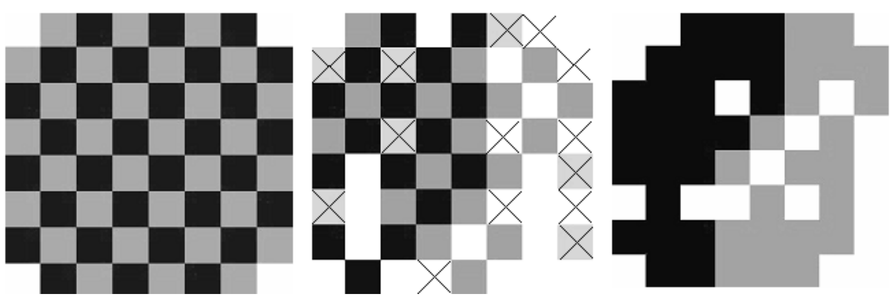
Formally, in multi-agent systems, emergence is a central property of dynamic systems based upon interacting autonomous entities (the agents). The knowledge of entities' attributes and rules is not sufficient to predict the behaviour of the whole system. Such a phenomenon results from the confrontation of the entities within a specific *structure of interaction*. That is, better knowledge of the generic properties of the interaction structures would make it easier to have better knowledge of the emergence process (ie. *morphogenetic dynamics*). From this point of view, to denote a phenomenon as "emergent" does not mean that it is impossible to explain or to model the related phenomenon. For this reason Epstein J.M. (1999) uses the word "generative" instead of "emergent" in order to avoid a philosophical debate about emergence.

Various attempts have been made to define emergence in an "objective" way. Some definitions refer to self-organisation (Varela et al., 1991), to entropy changes (Kauffman, 1990), to non-linearity (Langton, 1990), to deviations from predicted behaviour (Rosen, 1985, Cariani 1991) or from symmetry (Palmer, 1989). Other definitions are closely related to the concept of complexity (Bonabeau et al., 1995a, 1995b; Cariani, 1991; Kampis, 1991). In statistical physics (Galam, 2004), as well as for models in economics or social sciences explicitely based upon theses models (see for instance Durlauf, 1997, 2001 and the pioneering work of Galam et al, 1982 among others), emergence may be related with an *order parameter* which discriminates between at least two phases, each one with a different symmetry associated respectively to a zero and non-zero value of the order parameter. Each problem has its specific order parameter. For instance in the Ising model, where individual spins can takes the value $\{-1, +1\}$, the order parameter is the magnetization M, given by the sum of all the spin values divided by their total number. When $M = 0$, the state is paramagnetic, i.e. disordered in the spin orientations, while long range order appears as soon as $M \neq 0$. A majority of spins are then oriented along either -1 or $+1$, and an order is likely to emerge. Two ordered phases are thus possible in principle, but only one is effectively achieved. The order parameter provides a ”signature” for the emergent phenomenon. Although these definitions make use of concepts borrowed from physics and information science, they all involve inherently contingent aspects, as the presence of an external observer seems unavoidable. Even a change in entropy supposes that an observer be able to assess the probability of various states.

The *unavoidable presence of an observer* does not preclude, however, the possibility of extending the definition of emergence to include non-human observers or observers that are involved in the emerging phenomenon. In our quest for ”strong emergence”, we wish to assign the role of the observer to elements of the system itself, as when individuals become aware of phenomena affecting the whole society. This kind of self-observation is only possible because what is observed is a simplified state of the system. *Emergence deals* precisely *with simplification*.

2.2 What does emerge in Schelling’s model of spatial segregation?

Schelling’s model of spatial segregation (Schelling, 1969, 1971, 1978) is a pioneering example of an emerging phenomenon resulting from social interaction. Schelling’s aim was to explain how segregationist residential structures could spontaneously occur, even when people are not so very segregationist themselves. The absence of a global notion of segregationist structures (like the notion of ghettos) in the agent’s attributes (preferences) is a crucial feature of this model. Agents do not choose between living or not living in a segregationist structure, but have only local preferences concerning their preferred neighbourhood. Moreover, people have only weak segregationist behaviour, but the play of interactions generates global segregation. In Schelling’s original model, agents were placed on a 8-by-8 chessboard as shown in Figure 1 (Java applet).



(1-a) fully integrated population equilibrium (1-b) discontented agents are crossed (1-c) convergence after 4 iterations
Source: http://perso.univ-rennes1.fr/denis.phan/complexe/schelling.html and Phan, 2004

Fig. 1. Original (checkerboard) Schelling Model

Taking the “colour“ of agents as criterion for discrimination, agents choose a location where to live, depending on their individual tolerance threshold of different colours in their neighbourhood. Agents interact only locally with their 8 direct neighbours (within a so-called ”Moore” Neighbourhood). No global representation about the residential structure is available to them. Though agents may be weakly segregationist (each agent would stay in a neighbourhood with up to 62.5% of people with another colour), segregation occurs. Schelling used the following rule: an agent with one or two neighbours will try to move unless one of the two neighbours has the same colour as its own (which means a local tolerance of 50%); an agent with three to five neighbours requires at least two agents of same colour to stay (that is 33%, 50% and 60% local tolerance), and one with six to eight neighbours will stay if at least three of them are of the same colour (50%, 57,1%, 62,5% local tolerance).

Under Schelling’s behavioural assumption, a ”fully integrated structure” (Figure 1-a) is an equilibrium (an order) because no agent wishes to move. A ”fully integrated structure” is a structural pattern in which agents’ colours alternate in all directions. Because of border effects, no agent is located in the corners. The ”fully integrated structure” is an unstable equilibrium. A slight perturbation is sufficient to induce a *chain reaction* and the emergence of local segregationist patterns. In his example, Schelling extracted twenty agents at random, and added five at random in the free spaces. The discontented agents (crossed in Figure 1-b) move at random towards a new location in agreement with their preferences. These moves generate new discontented agents by a chain reaction until a new equilibrium is reached. In such equilibrium, local segregationist patterns appear, like in Figure 1-c.

Local interactions are sufficient for spatial homogeneous patterns to occur; spatial segregation is an emerging property of the system’s dynamics, while not being an attribute of the individual agents. Sometimes, local integrated (non-homogeneous) patterns may survive in some niches. But such integrated structures are easily perturbed by random changes, while homogeneous structures are more stable (frozen zones). Complementary theoretical developments on Schelling’s model of segregation can be found in the growing literature on this subject, for instance among economists like Zhang (2004a-b), Pancs, Vriend (2003, 2004), or sociologists like Broch, Mare (2004). Examples of advances in empirical investigations can be found in Clark (1991), Sethi Somanathan (2001), Koeler, Svoretz (2002), and experimentations in Ruoff, Schneider (2004). Our aim in this paper is to address emergent phenomena, as instantiated in Schelling’s model, in a new way. Emergence is currently debated for its cognitive and sociological aspects, from ontological and epistemic perspectives, in relation with the modern philosophy of mind (for a selection of papers, see for instance Intellectica 1998, and Gilbert, 1995, for links with sociology). There is also a debate within the artificial intelligence, artificial life and artificial society fields (see Gilbert,Conte,1995 in this later field). Entering or even summarizing those debates would fall outside the scope of the present paper. But some fundamental questions are worth asking about knowing in what way emergence occurs in Shelling’s model. Who is the observer? What does the higher level of organisation consist in? For whom does this level make sense? (Figure 2)

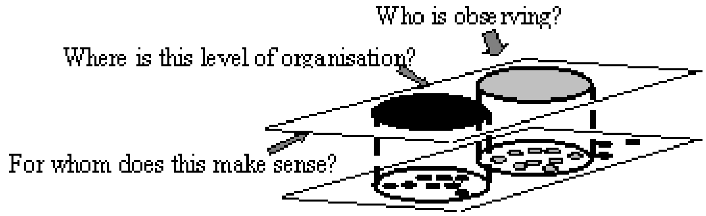


Fig. 2. Questions of emergence

In various definitions of emergence, the presence of an external observer seems unavoidable. Levels of organisation depend on an observer being able to discern subparts in the system and appropriate relations between them. There is no consistent way to say that some new phenomenon occurs at a higher level, be it some new form of operational closure or any form of deviation from expected behaviour, unless there was some pre-existing way to observe that higher level.

3 Emergence in ACE : from case studies to a formal definition

The first and second subsections provide two definitions coherent both with the design of multi-agent systems used in Agent Based Computational economics (Tesfatsion, 2002, Phan, 2004) and with important related features, like cognitive hierarchy, detection, and complexity. The first one (from Bonabeau, Dessalles 1997) defines the emergence as an unexpected complexity drop in the description of the system by a certain type of observer. The second one (from Muller 2002) defines emergence as a phenomenon observed at the interface of description levels. The latter definition introduces a useful distinction between ”weak” and ”strong” emergence.

3.1 Emergence as a complexity drop

In (Bonabeau, Dessalles, 1997), emergence is defined as an unexpected complexity drop in the description of the system by a certain type of observer. Such a definition is claimed to subsume previous definitions of emergence, both structural (dealing with levels of organisation) and epistemological (dealing with deviation from some model's predictions). In each case, the observer is able to detect a structure, such as the presence of relations holding between parts of the system, or some form of behaviour like a characteristic trajectory. Structural emergence occurs whenever the system turns out to be more structured than anticipated. This augmentation of structure can be characterised by a decrease of complexity.

$$E = C_{exp} - C_{obs}$$

Here, E stands for the amplitude of the emergence, C_{exp} is the expected structural complexity and C_{obs} the structural complexity actually observed. Structural complexity is defined as the algorithmic complexity relative to a given set of structural descriptors.

Algorithmic complexity, as defined by Kolmogorov, Chaitin and Solomonov, is defined by the shortest description that can be given of the system using a Turing Machine (Li, Vitanyi, 1993). This definition is sometimes considered of little use for finite systems, as the set of all systems of same size can be ordered; since each of these system can be characterised by its rank, nothing prevents the actual system to appear as the simplest one if it happens to be number one. In order to use algorithmic complexity to describe finite system, we abandon the generality of Turing machines, considering that the description tools are imposed by the observer. We define the relative algorithmic complexity (RAC) of a system as the complexity of the shortest description that a given observer can give of the system, relative to the description tools available to that observer. Emergence occurs when RAC abruptly drops down by a significant amount.

For our purpose here, we must restrict the definition. We consider a specific class of observers, in order to get closer to what human observers would consider as emergence. Following (Leyton 2001), we impose the observer's description tools to be structured as mathematical groups. In other words, any level of organisation that can be observed has operational closure and is structured as a group, and the only structures that can be observed are the invariant of a group of operations. Moreover, the observer is supposed to have hierarchical detection capabilities. This means that all elements of the system that the observer can consider have themselves a group structure. The observer may be considered as being a 'Leyton machine', for which any structure is obtained through a group-transfer of other structures (Leyton 2001).

Let us illustrate how emergence results from a complexity drop in Schelling's model. In a first stage, the external observer reconstructs the system by transferring (in the Leyton sense) one abstract inhabitant to form the entire population. The transfer group, in this case, is the group of 2-D translations. The operation is costly in terms of complexity, as each individual translation has to be instantiated. Then each abstract inhabitant is assigned a colour. This latter operation can be achieved through a transfer by the binary group Z/2Z. In a second stage, the external observer is now able to detect homogeneous clusters. She reconstructs the system in a different way. One first abstract cluster is obtained by translating one abstract inhabitant, as previously. Then this first cluster is itself translated to give the whole set of clusters. Finally, clusters are assigned colours through the binary group. Emergence, in this example, comes from the fact that the second construct is significantly simpler than the first one. The reason is that there are less colour assignments: only one per cluster instead as one per inhabitant. A crucial requirement for the emergence to be noticeable is that the shape of clusters be simple. For the system to be fully instantiated, the second construct must reshape the limits of each cluster through various groups of geometrical transformations. If there were no colours, or if the clusters had random shapes, there would be no gain in complexity. Conversely, emergence would be maximum in the extreme case in which all clusters had identical shapes, e.g. if they were square blocks.

Each transfer group can be seen as an organisation level. In Schelling's model, there could be more levels of organisation, for instance if clusters were arranged in a chessboard-like pattern. The Leytonian construct would be different and less complex in this case: the first cluster would be assigned a colour, then it would be duplicated through a binary symmetry group operating in colour space, then the couple would be transferred through the group of integer translations of the plan.

For structural emergence to occur, it is important that there be an unexpected complexity decrease. This may happen either because the higher structure detection was delayed, as when you take time to recognise a Dalmatian dog in a pattern of black and white spots. It may also happen when adding a new observable, instead of increasing the overall complexity of the system for the observer, paradoxically deceases it (Bonabeau, Dessalles, 1997). This latter case is well illustrated by our extension of Axtell et al.'s experiment (see Phan, Galam, Dessalles, 2005).

3.2 Emergence occurring in a system with several levels

Following Forrest's definition of emergent calculation (Forrest 1990), Müller (2002) defines emergence in SMA as occurring between two organisation levels, distinguishing the process and the observation of that process. The process concerns the evolution of a system formed by entities in interactions. These interactions may generate *observable epiphenomena*. At the observation level, *epiphenomena are interpreted as emerging through specific calculation* (i.e. like order parameter). For Müller, "*weak emergence*" arises when the observer is external to the system, while "*strong emergence*" arises when the agents involved in the emerging phenomenon are able to perceive it. In this later configuration the identification of epiphenomena by the agents in interaction in the system will involve a *feedback from the observation to the process*. There is a coupling between the process level and the observation level by the way of the agents. Emergence is thus *immanent* in such a system.

More specifically, for Müller, a phenomenon is emergent if:

- (A) There is a system composed of agents in interaction with each other and with their environment. The description of this system as a process is formalized in a language D
- (B) The dynamics of this system produces a structural phenomenon observable in the "traces of execution"
- (C) The global phenomenon is observed by an external observer (weak emergence) or by the agents themselves (strong emergence) and is described in a language distinct from D.

When compared with Forrest's definition (Forrest 1990), Müller's definition presupposes the existence of two languages of description, which are distinct according to the level considered. This distinction only materializes the presence of levels already hypothesised by Forrest. On the other hand, it is interesting to note that Müller distinguishes the system formed by the interacting agents from the process that governs their behaviour. This enables him to choose the position of the level of observation with respect to the agents. Müller's contribution lies then mainly in the distinction between two categories of emergence according to the position of the level of observation w.r.t. process. In strong emergence, agents are observers themselves, this de facto entails a feedback loop between the micro (agent based) level of observation and the macro level of the process. In weak emergence, the observer is external with the process and there is no necessarily coupling.

Müller illustrates weak emergence by means of the example of foraging ants which move between their nest and a food source. Each ant deposits on its passage some traces of pheromone which attract the other ants, and create an interaction between them (1). These interactions build a stable and observable phenomenon (2). An external observer may interpret this phenomenon as a ”path”. Moreover, the accumulation phenomenon based on interaction drives the ant colony to find the shortest path between their nest and a source of food. Emergence is weak because the dynamics depends only on the traces of pheromone (1-2) and not on the qualification of these traces as an ”shortest or optimal path”, which does not exist in the ants' head.

The category of strong emergence is important for to model artificial societies (Gilbert, 1995). Indeed, the reflexivity meditated by the agents' ”consciousness” appears to be a determinant characteristic that distinguishes systems involving human agents from systems made of non conscious or material entities.

In Schelling's model, there would be strong emergence if agents, rather than merely sampling neighbouring densities, were able to perceive forming homogeneous clusters in the town and if their perception could affect their decisions. Strong emergence is particularly important in economic modelling, because the behaviour of agents may be recursively influenced by their perception of emerging properties. Emerging phenomena in a population of agents are expected to be richer and more complex when agents have enough cognitive abilities to perceive the emergent patterns. Such feedback loops between emerging collective patterns and their cognitive components clearly occur among agents in human societies. They may obey laws that are still to be understood. Our aim here is to design a minimal setting in which this kind of strong emergence unambiguously takes place.

To summarize, if there is *strong emergence* in the sense of Müller, the system becomes reflexive, through the mediation of the agents. (A) Agents are equipped with the capacity to observe and to identify an epiphenomenon in the process which represents the evolution of the system in which they interact. This capacity of observation and the field of such observation must then be sufficiently broad to encompass the phenomenon as a global one. (B) The agents can describe this epiphenomenon in a ”language” other than that which is used to describe the process (C) The identification of an ”emergent” epiphenomenon by the agents involves a change of behaviour, therefore a feedback of the level of observation on the process

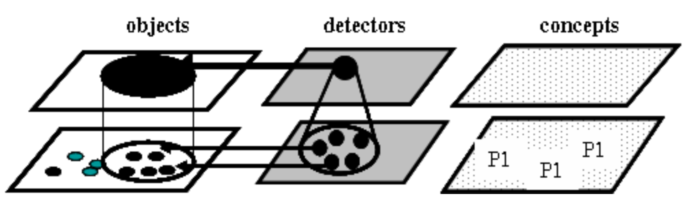


Fig. 3. Parallelism between hierarchies : description, observations and conceptual level

Emergent phenomena are naturally described in a two-level architecture (Figure 3). In such a framework, objects at the two levels only exist because some observer is able to detect them. The detected object at the upper level is composed by objects of the first level. Correspondingly, the upper level detector is triggered by the activity of lower level detectors. The system's complexity, defined as the minimal description that can be given of its state, drops down by a significant amount when an upper-level detector becomes active, as its activity subsumes the activity of several lower-level detectors.

According to this point of view, one can reinterpret the Müller's definition using a distinction due to Searle (1995) between entities that are independent of the observer (the process and the phenomena which results from it) and entities that occur within the observer (identification and interpretation of an epiphenomenon). According to this interpretation, *emergence becomes a category relative to an observer*, and in the case of a human observer (or an agent supposed to be represented an human), a subjective category. Note that Müller's definitions and the above definition of structural emergence as complexity drop are compatible. Müller's distinction between two description languages presupposes that the upper language, available to the observer, provides it with a simpler description of the epiphenomena than what was available at the process level.

3.3 Learning and ”intrinsic emergence”

Crutchfield (1994) Bersini (2004), Philemotte, Bersini (2005) propose to consider an alternative definition of emergence, called 'intrinsic emergence'. They suggest to characterise emergence as an autonomous increase in the system's computational capabilities. Such a definition is supposed to be more 'objective' as a natural way to avoid the presence of an external observer in charge of detecting emergence. Philemotte, Bersini (2005) implemented a situation of intrinsic emergence. In their system, a cellular automaton is evolved through a genetic algorithm (GA) until it is able to perform some arithmetic operations on a limited set of operands. As usual for cellular automatons, the rules which, for each cell, decide of its next state, take as input the previous state of neighbouring cells. In Philemotte, Bersini's system, a second genetic algorithm is in charge of filtering inputs for the cellular automaton, so as to make the learning task easier for the first GA. Intrinsic emergence is claimed to occur whenever the second GA is able to isolate a relevant portion of the neighbouring input and thus to significantly improve the learning efficiency of the overall system. Philemotte, Bersini were able to observe such sudden improvements when both genetic algorithms cooperate.

This definition is original, and somewhat differs from the previous ones, which were limited to the description of structural patterns. We may call it behavioural emergence, as the criterion for emergence is a discontinuity in performance rather than a discontinuity in structural complexity. We may, however, ask what is emerging. If the general definition of intrinsic emergence is restricted to describe some discontinuity in efficiency, then the answer is that nothing does emerge. In Philemotte, Bersini's experiment, however, a relevant input filter can be said to emerge. For some definition of complexity, indeed, intrinsic emergence is well described by definition (1). The measure of complexity to be considered here is the size of the relevant search space. When systematically ignoring a portion of the input, the second GA dramatically reduces the space where the first GA will find an efficient rule for the cellular automaton. This presupposes, however, that the input filter does not exclude convenient solutions. If complexity is set to a maximal value when no adequate rule is learned, then intrinsic emergence can be said to correspond to a complexity drop. Note, however, that intrinsic emergence, contrary to structural emergence, does not rely on the complexity of structure, e.g. the complexity of hierarchical group structure, but relies on learning efficiency which directly correlates with the size of the filtered search space.

4 - The emergence of classes in a population game: overview of the original model and discussion

In order to illustrate definitions introduced in previous section, our aim is to design a model that gives rise to ”strong emergence”. In this paper, we provide an overview of the original model and discuss the conditions of implementation. Results of the implementation in a multi-agent framework of multi-level strong emergence with a detection process are presented in a companion paper (Phan, Galam, Dessalles, 2005). We start from a model of class emergence (Axtell, et al. 2001) in which agents play a population game and tend to correlate other players’ behaviour with fortuitous visible but meaningless characteristics (tags). This model is a nice example of emergence in an ACE population game. The authors show how intrinsically meaningless ”tags” associated with agents can acquire social salience over time such that tag-based classes emerge at the social level from the decentralised interactions of many individuals who accumulate over time information about the behaviour of others by the way of history - based expectations. On some occasions, these fortuitous tags turn out to be reliable indicators of dominant and submissive behaviour in an iterative Nash bargaining tournament.

4.1 Axtell, Epstein and Young’s model of class emergence

This model is a ”random pair wise” type of population game (Young 1998, Blume, 1997) with linear trembling hand. That is, Nash equilibrium can be reached without any assumption about common knowledge. Early analytical results can be found in Young (1993). During the game, agents are randomly paired and at each time step play a ”one-shot ” game with their opponent. Agents choose the strategy which is their best response given their beliefs (a ” mixed strategy ”) about the behaviour of the others, drawn by induction from a distribution of observed strategies kept in a finite memory of size m. At each time steps, agents change partners and actualise their beliefs depending on the result of the last meeting. Agents have a linear positive probability of deviation (trembling hand). The formal context is thus stochastic and the concept of stability used by the authors is due to Foster, Young (1990) for stochastic evolutionary games. The one-shot negotiation between pairs of agents is drawn from the one step Nash bargaining model. That is, each player tries to share a "cake" of size 100 with its opponent by opting for one of three possible strategies: "High " (H), "Medium (egalitarian) " (M), and " Low" (L). The corresponding percentages of the "cake" claimed by players can be fixed, without loss of generality, to 70

Table 1. best reply equivalent matrix for a bilateral game of agent i

	$S_1 : H$	$S_2 : M$	$S_3 : L$
$S_1 : H$	$(0, 0)$	$(0, 0)$	$(70, 30)$
$S_2 : M$	$(0, 0)$	$(50, 50)$	$(70, 30)$
$S_3 : L$	$(30, 70)$	$(30, 50)$	$(30, 30)$

Player i in rows / Player k in columns

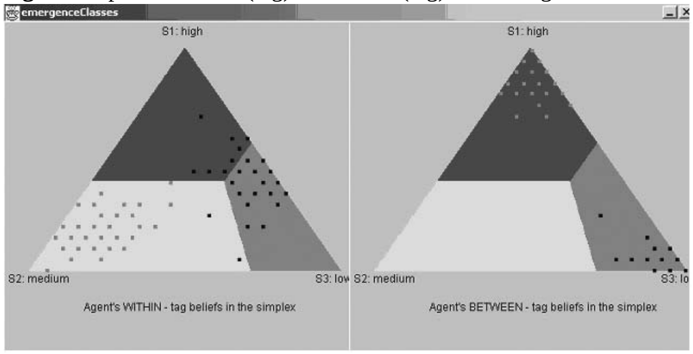
The authors distinguish three situations. First, there are tree situations where the agents' payoff is null and consequently highly inefficient (because they proposed more than 100

The strategies played within the population are triplets $\sigma_{im} = (p, q, 1 - p - q)$, and each agent infers its expected payoff from a historical sample of size m, say: $\sigma_{im} = (p_i, q_i, 1 - p_i - q_i)$. At each time step, randomly paired agents play their best response against their own expected mixed strategy im with probability $(1 - \epsilon)$ and play at random with probability ϵ (linear trembling hand " *la Young*"). When playing, an agent observe the strategy of its opponent and updates its belief by removing from its memory the oldest value and by updating its list by inserting the last strategy observed. The state of belief of an agent can be represented by a point on a simplex of size 2 used to represent the expected mixed strategies of this game (Figure 3).

The initial beliefs can be initialized in a random way, or in a targeted zone. But an initial form of heterogeneity of the beliefs is necessary to usefully explore the dynamic properties of this model. Indeed let us suppose that all the agents initially form the belief that their opponents play M. Their best reply, conditionally to this initial belief will precisely consist in playing M, which will reinforce the overall initial belief of these agents. The initial beliefs of an agent can be interpreted as their "cultural" heritage and their updated beliefs as the product of "the history" of their last meetings (an "historical" form of interactional heterogeneity, since agents' history differ). Let us note that in the AEY's model, the agents do not have common belief nor beliefs upon the beliefs of other agents, but only about the distribution of strategies. When the beliefs of the agents im are located in the same zone of the simplex (say by example, "M"), they are in a weak sense "shared beliefs", because their best response is the same "M", but this situation is not recursive and agents have not common beliefs. Finally, using the results of Young (1993), the authors show that the only stochastically stable solution corresponds to a situation where almost all the agents play "M".

In a second time, the authors introduce two types of agents, differentiated by an observable external sign (a tag) which enables them to be identified (grey and black in our simulation). The authors assume that this sign does not have any intrinsic significance (”completely meaningless”).

Fig. 4. Simplexes Within (tag) - Between (tag) and emergence of classes



Axtell et al. (2001) model, simulated on Moduleco-MadKit

However, the agents memorize the sign of the opponents whom they met and calculate the average behaviour corresponding to each type. There are thus two groups, determined beforehand by the tags, but this is not sufficient to cause a differentiated behaviour sensitive to tags, which could result in a shared belief on the behaviour of the members of these groups. However, in this model with two tag types, beliefs about the opponent’s strategy may diverge depending on the opponent’s tag, leading to between-types (grey against black, Figure 3, right) and within-type (grey against grey or black against black, Figure 3, left) responses. By definition, the formation of ”classes” corresponds to the relative stabilization of distinct beliefs based on the group, leading to an equitable intra-group behaviour (within), and an unfair share between classes (the opposite case exists, but can be regarded as ”pathological”). In the situation displayed on Figure 3 left, grey dots show equitable behaviour (they play M) when encountering agents of their kind (within = intra-group), whereas blacks dots don’t (but they have moved close to the zone of equity). The situation displayed on the right shows that blacks dots have the belief that grey dots adopt in majority a ”dominated” behaviour (L) and their best response thus consists in claiming a large share (H). Conversely, grey agents have the belief that black agents preferentially show a ”dominating” behaviour (H) and their best response then consists in adopting a dominated attitude by accepting a small share (L). Therefore, both beliefs reinforce each other.

In the model with tags, (as in the case without tags), the stochastic process is ergodic and the only stable regime is the ”equitable” one: MM. More specifically, if the length m of the agents’ memory and the ratio of the number of agents N to this length (N/m) are ”sufficiently large” while the trembling hand effect remains ”sufficiently weak”, the ergodic (invariant) probability to be in the ”equitable” area is high. However, if m is large and N small, the inertia of the system, (i.e. the time before reaching or leaving an area) can be very important (”broken ergodicity”). This is true in particular for the transition from the mode ”with class” towards the equitable standard.

4.2 Discussing the emergence in Axtell et al. (2001)

One limit of Axtell et al. model is that dominant and submissive classes remain implicit within the system. Agents are designed to take biased decisions depending on tags, but any actual bias resulting in a divergent attitude toward different tag bearers remains contingent and is never represented as such in the system. As a consequence, behaviour classes only emerge in the eye of external observers.

Can we say that structural emergence takes place in Axtell et al.'s model?

From an external observer perspective, the expected situation is the initial one, when every agent adopts its own strategy. Its complexity is maximal, as a description of the situation requires each agent to be assigned a location on the simplex. In this context, taking displayed tags into account is expected to bring supplementary complexity, as it requires additional instantiation. Paradoxically, a less complex description of the system may be achieved by the detour through the tag. Once the population self-organises in two tag-consistent clusters, there are few deviations left to instantiate whenever the tag happens to be a good predictor of behaviour. In Leyton's terms, when the initial abstract cluster is transferred to give the two classes, behaviour is assigned to clusters simultaneously with tag value, instead of requiring independent instantiations for each individual.

Emergence, here, results from an unexpected decrease of complexity, in conformity with definition (1). However, it cannot be considered a case of strong emergence, as individual agents have no way to observe it. For strong emergence to occur, the capabilities of agents have to be extended. From the agents' perspective, the problem is to predict the strategy of their next partner. As there are three possible strategies, the agent's inference process is to partition a sample of the population, constituted through random encounters, into three clusters to build a mathematical expectation estimator. When the agent does not pay attention to tags, its sample of the population would appear complex to it. If the three different observed strategies are equally represented in the agent's sample, then the structure is maximally complex. When one behaviour predominates, then the situation may appear slightly less complex, as it can be described by first assuming that all individual belong by default to the majority class, and then by accounting for deviations.

In Axtell et al.'s model, agents are not equipped with the ability to assess structural complexity. Moreover, in the model with tags, they have a built-in bias that forces them to split their sample of the population into two separate classes according to the displayed tags. Things would be different if agents had the possibility to decide whether to pay attention to tags or not. To extend the model, we make the additional hypothesis that the agents are more cognitive than in the basic model, i.e. they seek the best way to sort out their sample of the population according to observed behaviours. To do this, they may rely on various cognitive (classification) rules:

- C1- Maintain three separate lists and distribute the m individuals of the sample among these lists according to their behaviour.
- C2- Maintain only two lists, considering the majority behaviour to be the default

- C3- Constitute a tag-behaviour association matrix, and apply cognitive rule C1 for deviant individuals only.

When behaviours are not evenly distributed, C2 is clearly less complex than C1. When tags are irrelevant, C3 is more complex than the two others. When tags become relevant, however, C3 may become the least complex rule. As agents themselves may notice a complexity drop, this would be a case of strong emergence.

The main point of Axtell et al.'s model is that the predictive value of tags is not due to sole random drift. There may be positive reinforcement between the classification rule of agents and their actual behaviour. However, this positive feedback is an automatic consequence of the built-in bias that prompts agents to make two classes according to the binary label. The reason is that all agents make the same simplifying assumption that there are two classes, plus the assumption that there may be a connection between class and behaviour.

A natural extension of Axtell et al.'s model is to replace the built-in simplifying assumption about the pre-existence of two classes by a mere bias favouring simplicity. To do so, let us reason about a more general setting in which T different binary tags may be displayed. Without any bias, cognitive rule C3 is intractable. There are 2T different tag combinations, and thus 32T different tag-behaviour association matrices, which is a huge search space as soon as T has a significant value. A first simplifying bias consists in considering that classes must be characterised by a conjunction of some of the tags. The size of the search space for learning the classes is now (3T)3, as for each behaviour class each tag may be specified either to 0 or to 1 or remain unspecified. For T=5 for instance, the search space is reduced from 1015 down to 107. Searching through spaces of such significant size may prove highly inefficient without any additional bias, especially if the decisions made by the agents are supposed to have some impact on their fate. A further and natural bias consists in having agents examine simpler hypotheses first. Tag combinations may be ranked according to their complexity, e.g. according to the number of specified values. Following a bias toward simplicity, agents will consider elementary tags first, then combinations of two tags, and so on.

The crucial aspect of this extension of Axtell et al.'s model is that agents rely on the same set of tags and on the same simplifying bias (the ordering of tags may be, of course, agent-specific). This hypothesis may be the key to keep the positive reinforcement between classification and actual behaviour valid. Moreover, it avoids considering built-in classes, as in the initial model. Agents must figure out for themselves what characterises the behaviour classes, before applying cognitive rule C3. Note that the possibility of tag combinations allows naturally for minority classes (in an evenly distributed population, 25

If we abandon the oversimplifying assumption that there are two predefined classes, then a learning rule to extract the relevant tags is necessary. If we make the additional hypothesis that agents prefer simple (costless) classifications, then not only do we have a natural extension of the initial model. Such a system is also rich enough to produce strong emergence. Because classes must be learned, agents have to decide when to shift from cognitive rule C2 to cognitive rule C3, and the criterion to do so is a complexity drop. The more agents apply C3 to a given tag combination, the lower the complexity through positive reinforcement of classes, and the greater the emergence. And this emergence serves as input for more agents to adopt C3.

4.3 Implementation and results

Forthcoming ../..

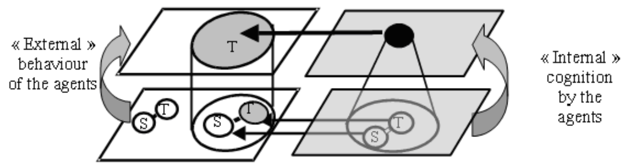


Fig. 5. Parallelism between hierarchies : description, observations and conceptual level

References

1. Axtell R. Epstein J.M., Young H.P. (2001) The Emergence of Classes in a Multi-agent Bargaining Model in Durlauf, Young eds.,Social dynamics, The MIT Press, p. 191-212
2. Baas, N.A. 1994. "Emergence, Hierarchies, and Hyperstructures," pp. 515-537, in C. G. Langton, ed. A life III, Santa Fe Studies in the Sciences of Complexity, Proc. Volume XVII, Redwood City, CA: Addison-Wesley.
3. Bedau, M.A. (1997) Weak Emergence, Nos, Vol. 31, Supplement: Philosophical Perspectives, 11, Mind, Causation, and World, p. 375-399
4. Bedau, M. A.((2002) Downward causation and the autonomy of weak emergence, Principia 6-1 June, special issue on Emergence and Downward Causation, p. 5-50
5. Bersini, H (2004) Whatever emerges should be intrinsically useful.in: Artificial life 9, The MIT Press p. 226-231
6. Blume L. (1997) Population games, in Arthur, Durlauf, Lane eds. p.425-460
7. Bonabeau E., Dessalles J-L. (1997) Detection and Emergence, Intellectica, 25:89-94.
8. Bonabeau, E., Dessalles, J.-L. and Grumbach, A. (1995a) Characterizing emergent phenomena (1): A critical review. Rev. Int. Syst. 9:327-346.
9. Bonabeau, E., Dessalles, J.-L. and Grumbach, A. (1995b) Characterizing emergent phenomena (2): a conceptual framework. Rev. Int. Syst. 9:347-369.
10. Broch, Mare (2003) Neighborhood Choice and Neighborhood Change California Center for Population Research On-Line Working Paper Series CCPR-003-03, June.
11. Bourgine P., Nadal J.P. eds. (2004) Cognitive Economics, Spinger Verlag
12. Durlauf S.N., Young P. (2001) eds., Social dynamics,The MIT Press, Cambridge Ma.
13. Epstein J.M. (1999) Agent-Based Computational Models and Generative Social Science, Complexity, Vol.4-5:41-60.
14. Campbell DT (1974), Downward causation, in hierarchically organized biological systems. In: Studies in the Philosophy of Biology: Reduction and Related Problems, Ayala, Dobzhansky, eds. Macmillan, London
15. Cariani, P. (1991) Adaptivity and emergence in organisms and devices, World Futures 31, pp. 49-70, Gordon & Breach Science Publishers

16. Clark W.A.V. (1991) Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model, Demography 28, P. 1-19.
17. Crutchfield (1994) Is anything ever new? considering emergence. In G. Cowan, D.P. Melzner, D., eds.: Integrative Themes. Volume XIX of Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley Publishing Company, Reading, Massachusetts
18. Davidson, D. (1970) Mental Events, in Lawrence Foster and J. W. Swanson, eds., Experience and Theory, Duckworth, London.
19. Davidson D. (1980) Actions and Events, Clarendon, Oxford
20. Durlauf S.N. (1997) Statistical Mechanics Approaches to Socioeconomic Behavior, in Arthur, Durlauf, Lane (eds), *The Economy as an Evolving Complex System II*, Santa Fe Institute Studies in the Sciences of Complexity, Volume XXVII, Addison-Wesley.
21. Durlauf.S.N. (2001) A Framework for the Study of Individual Behavior and Social Interactions, Sociological Methodology Volume 31 Issue 1, 1 January pp. 47-87
22. Emmeche, C., Kppe, S. & Stjernfelt, F. 1997. Explaining emergence-towards an ontology of levels, Journal for General Philosophy of Science, 28: 83-119.
23. Epstein J.M. (1999) Agent-Based Computational Models and Generative Social Science, Complexity, Vol.4-5:41-60.
24. Ferber, J., Labbani O., Müller J.-P., Bourjault. A. (1997) Formalizing emergent collective behaviours: Preliminary report. International Workshop on Decentralized Intelligent and Multi-Agent Systems, St. Petersburg, p. 113-122.
25. Ferber J. (1999) Multi-agent Systems, Addison Wesley Reading, MA.
26. Forrest S. (1990) Emergent Computation : Self-organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks, in Emergent computation, MIT Press, Cambridge, M.A., 1990, p. 1-11.
27. Freidlin, M., Wentzell A. (1984) Stochastic Dynamical Systems Theory, Springer Verlag, NY
28. Foster, D.P., Young H.P. (1990) Stochastic Evolutionary Game Dynamics, Theoretical Population Biology 38:219-232.
29. Galam S., Gefen Y. , ShapirY. (1982) Sociophysics: a new approach of sociological collective behavior, Journal of Mathematical Sociology **9**, 1
30. Gilbert N. (1995) Emergence in social simulations, in Gilbert, Conte. eds. 114-156.
31. Gilbert N., Conte R. eds. (1995) Artificial Societies : The Computer Simulation of Social Life UCL Press London.
32. Gillett C. (2002a) The Varieties of Emergence: Their Purposes, Obligations and Importance, Grazer Philosophische Studien, vol.65, pp.89-115.
33. Gillett C. (2002b) Strong Emergence as a Defense of Non-Reductive Physicalism: A Physicalist Metaphysics for 'Downward' Determination, Principia 6-1 June, special issue on Emergence and Downward Causation,, p.83-114.
34. Granovetter M.S. (1985) Economic Action and Social Structure: The problem of embeddedness, American Journal of Sociology, 91 p. 481-510.
35. Holland, J.H. (1998) Emergence From Chaos to order, Helix Books Addison-Wesley pub. Readings, Mass.
36. Kampis, G. (1991) Self-Modifying Systems in Biology and Cognitive Science Pergamon
37. Kim, J. (1992) Downward causation, in emergentism and nonreductive materialism, in Beckermann, Flohr, Kim, eds., Emergence or Reduction? Walter de Gruyter.
38. Kim, J. (1993) Supervenience and Mind, Cambridge: Cambridge University Press.
39. Kim, J. (1995) Emergent properties, in T. Honderich, ed. The Oxford Companion to Philosophy, Oxford University Press p. 224.
40. Kim, J. (1999) Making sense of Emergence, Philosophical Studies, 95 p.3-36.

41. Koehler G. Skvoretz J. (2002) Preferences and Residential Segregation: A Case Study of University Housing WP
42. Langton C. (1989) Computations at the edge of chaos, Physica D, 42
43. Li M. Vitanyi P.M.B. (1993) An Introduction to Kolmogorov Complexity and Its Applications. Springer-Verlag, New York, Berlin
44. Lewes, G. H. (1874) in Emergence, Dictionnaire de la langue philosophique, Foulqui.
45. Leyton, M. (2001) A generative theory of shape, New York: Springer Verlag
46. McLaughlin B. P. (1992) The rise and fall of British emergentism, in Beckermann, Flohr, Kim, eds., Emergence or Reduction? Berlin: Walter de Gruyter.
47. McLaughlin B. P. (1997) Emergence and supervenience, Intellectica 25 p.25-43.
48. Mill, J.S. (1843) System of Logic, Longmans, Green, Reader, and Dyer, London.
49. Muller J.P. (2002) Des systèmes autonomes aux systèmes multi-agents : interactions, émergence et systèmes complexes Rapport de l'HDR, 8 novembre, 58p.
50. Nash, J. (1950) The Bargaining Problem, Econometrica 21:128-140.
51. Palmer, R. (1989) Broken Ergodicity, in Stein ed. Lectures in the Sciences of Complexity, Vol.1, SFI Studies in the Sciences of Complexity, Addison-Wesley, Reading, Mass.
52. Pancs, R., Vriend, N.J. (2003) Schelling's Spatial Proximity Model of Segregation Revisited, Dept. of Economics Working Paper No. 487 Queen Mary, University of London
53. Phan D. (2004) From Agent-Based Computational Economics towards Cognitive Economics, in Bourgine, Nadal, eds., Cognitive Economics, Spinger Verlag p. 371-398.
54. Phan D. (2004b) Hierarchy of cognitive interactive agents and statistical mechanics: how Object Oriented Programming highlight the connection, in Coelho H., Espinasse B. Seidel M-M.eds., 5th Workshop on Agent Based Simulation, Lisbon, Portugal, 3-5 May 2004 SCS Publishing House, Erlangen, San Diego, p. 69-76
55. Philemotte, Bersini (2005a) Intrinsic Emergence Boost Adaptive Capacity
56. Philemotte, Bersini (2005b) Co-evolution of Effective Observers and Observed Multi-Agents System
57. Rosen, R. (1978) Fundamentals of measurement and representation of natural systems, North Holland.
58. Ruoff G. Schneider G (2004) Segregation in the Class Room: An Empirical Test of the Schelling Model, meeting of the working group on decision theory of the German Political Science Association, University of Bamberg, July 9-10
59. Schelling T.S. (1969) Models of Segregation, American Economic Review, Papers and Proceedings, 59-2:488-493.
60. Schelling T. S.(1971) Dynamic Models of Segregation, Journal of Mathematical Sociology; 1, p.143-186.
61. Schelling T.S. (1978) Micromotives and Macrobehavior W.W. Norton and Co, N.Y.
62. Searle J.R. (1995) The Construction of Social Reality, Free Press
63. Sethi R. Somanathan R (2004), Inequality and Segregation, Journal of PoliticalEconomy p.112-6
64. Sperry, R. W.(1969) A modified concept of consciousness, Psychol. Rev. 76, 532-536
65. Sperry, R. W. (1986) Discussion: Macro- versus micro-determinism (A response to Klee), Philosophy of Science, 53(2), p.265-270.
66. Sperry, R. W. (1991) In defense of mentalism and emergent interaction, Journal of Mind and Behavior 12, 221-245
67. Tesfatsion L. (2002) Economic Agents and Markets as Emergent Phenomena, Proceedings of the National Academy of Sciences U.S.A., Vol. 99-S3:7191-7192
68. Tesfatsion L. (2002b) Agent-Based Computational Economics: Growing Economies from the Bottom Up, Artificial Life, Volume 8 - 1, MIT Press, p.55-82.

69. Van de Vijver G. (1997) Emergence et explication, Intellectica 25:7-23.
70. Varela F., Thompson E., Rosch E. (1991) The embodied mind, MIT Press.
71. Young H.P. (1993) An Evolutionary Model of Bargaining, JET 59-1:145-168.
72. Young H.P. (1998) Individual Strategy and Social Structure, Princeton University Press, New Jersey.
73. Zhang J. (2004a) Residential Segregation in an All-Integrationist World, Journal of Economic Behaviour & Organization, 54, p.533,550
74. Zhang J. (2004b) A Dynamic Model of Residential Segregation, Journal of Mathematical Sociology, 28, p.147-170