

# Agent Cognitive capabilities and Orders of Emergence: critical thresholds relevant to the simulation of social behaviours.

Chris Goldspink<sup>1</sup>, Robert Kay<sup>2</sup>

**Abstract** In this paper we provide a brief recount of alternative approaches to what we argue is a fundamental issue for our understanding of sociality – the micro-macro problem or, as we refer to it here, the problem of social emergence. We then discuss recent attempts to identify how the range and type of emergent phenomena changes as a result of changes in the fundamental characteristics of micro-agents. We conclude that there appear to be a number of critical thresholds, notably that which arises when agents become constitutively autonomous and subsequently also develop behavioural (sensori-motor) autonomy. It is the combination of these two levels of autonomy which accounts for what we typically call ‘cognition’ in biological agents. Current artificial intelligence models attempt to replicate the ability without autonomy. While this approach is being seen as increasingly problematic in robotics it appears yet to have influenced approaches to social simulation. We propose achieving behavioural autonomy as a goal and focal point for future simulation research. We argue that this is the minimum threshold needed to achieve social emergence. We illustrate this by discussing the concept of social ‘norm’ as an ‘attractor’ in a phenomenal domain of structurally coupled behaviour..<sup>12</sup>

## 1 INTRODUCTION

Building and working with artificial societies using the methods of multi-agent social simulation serves us in several ways – it allows us to operationalize social theories and to compare simulated behaviours with those observed in the real world and it allows us to build new theory by exploring the minimal mechanisms that might explain observed social behaviour. Most importantly, it provides a unique ability to explore the interplay between levels of phenomena and to understand dynamic properties of systems. A great deal can and has been achieved in both these areas with even the simple methods we currently have available. However, Keith Sawyer [1] has recently reminded us that, to date, we have worked with agents of very limited cognitive capability and that this necessarily limits the range and type of behaviour which can be explored. This echoes a sentiment made a decade ago by Christiano Castelfranchi [2] that social simulation is not really *social* until it can provide an adequate account of the implication

of the feedback between macro and micro which becomes possible with higher cognitive functioning of social agents.

This paper examines the relationship between agent capability and orders of emergence in order better to define the critical thresholds which limit our capacity to simulate certain classes of social phenomena.

In many respects, developments in our capacity to simulate artificial societies have led us to confront anew a long-standing issue within social theory. This problem is variously referred to as the micro-macro problem, the problem of structure and agency or social emergence. This problem has been a long term focus of our collaboration [see 3, 4]. Over the past decade we have worked towards a theory of sociality which can provide a coherent and consistent account of the interpenetration (circular causality) of micro and macro phenomena – i.e. which can provide a substantive account of fundamental social generative mechanisms. No such theory currently exists. This current paper is a continuation of that work but also has its origin in one author’s involvement with the EU funded project titled Emergence in the Loop (EMIL). Through EMIL we aim to a) provide a theoretical account of the mechanisms of normative self-regulation in a number of computer mediated communities b) specify the minimum cognitive processes agents require to behave in normative ways c) develop a simulator which can replicate the range and type of normative behaviour identified by the empirical research so as to further deepen our understanding of how and under what conditions normative self-regulation is possible.

## 2 A BRIEF RECOUNT OF THE PROBLEM

The notion of emergence has a long history. Unfortunately the concept remains ill defined ambiguous and contentious, leading to the criticism that it stands as little more than a covering concept – used when no adequate account or explanation exists for some unexpected phenomena. The origin of the concept has been attributed to George Henry Lewes, in 1875 [5]. It subsequently found wide adoption within the philosophy of science but has been advanced within four streams: *philosophy*, particularly of science and mind; *systems theory*, in particular complex systems; *social science* where it has largely been referred to under the heading of the micro-macro link and/or the problem of structure and agency; and more recently in theoretical biology, cognitive theory and robotics. Interestingly there has been relatively little cross influence between these streams.

---

1 Centre for Research in Social Simulation, Department of Sociology, University of Surrey, Guildford, GU1 7XH, UK, c.goldspink@surrey.ac.uk

2 Executive Director and co-founder of Incept Labs, Sydney, Australia, rkay@inceptlabs.com.au

## The Contribution from Philosophy of science

The philosophy of science and philosophy of mind stream is arguably the oldest – some date it back to Plato [6] but the debate is widely seen as having come to focus with the British Emergentists [7-9]. This school sought to deal with the apparent qualitatively distinct properties associated with different phenomena (physical, chemical, biological, mental) in the context of the debate between mechanism and vitalism. This stream remains focused on explaining different *properties* of classes of natural phenomena and with the relationship between brains and minds [See 10 for a recent summary of the positions]. As a consequence this has been the dominant stream within artificial intelligence. Peterson [6: 695] summarizes the widely agreed characteristics of emergent phenomena within this stream as follows. Emergent entities:

1. Are characterized by higher-order descriptions (i.e. form a *hierarchy*).
2. Obey higher order *laws*.
3. Are characterized by *unpredictable novelty*.
4. Are *composed of* lower level entities, but lower level entities are *insufficient* to fully account for emergent entities (*irreducibility*).
5. May be capable of *top-down causation*.
6. Are characterized by *multiple realization or wild disjunction* [11] (alternative micro-states may generate the same macro states).

Within this stream there is a concern with both upward and downward causation and it is the possibility for the later which attracts most argument. A key concept is *supervenience*: a specification of the ‘loose’ determinism held to apply between levels such that ‘...an entity cannot change at a higher level without also changing at a lower level’ [12: 556]. Advocates of supervenience argue that properties associated by emergent structures exist only due to the properties of the underlying constituents and, in having no unique causal power other than those derived from those constituents, comprise only epiphenomena – they are not ‘real’. This controversy persists within philosophical circles although it appears to derive in large part from an extreme form of physicalism [13]. Practicing physicists appear to have fewer problems with the concept than philosophers of mind. Physicists Clayton and Davies [10], for example, specify downward causation as involving macro structures placing *constraint* on lower level processes hence ‘*Emergent entities provide the context in which local, bottom up causation takes place and is made possible*’ [6: 697]. Davies [14] argues that the mechanism of downward causation can usefully be considered in terms of boundaries. Novelty, he argues, may have its origin in a system being ‘open’. He concludes:

... top-down talk refers not to vitalistic augmentation of known forces, but rather to the system harnessing existing forces for its own ends. The problem is to understand how this harnessing happens, not at the level of individual intermolecular interactions, but overall – as a coherent project. It appears that once a system is sufficiently complex, then new top down rules of causation emerge (Davies 2006: 48).

For Davies then, top-down causation is associated with self-organization and may undergo qualitative transitions with

increasing system complexity. For Davies also it is the ‘openness’ of some systems that ‘provides room’ for self-organizing process to arise, but he concludes, ‘*openness to the environment merely explains why there may be room for top-down causation; it tells us nothing about how that causation works.*’ The devil then, is in the detail of the mechanisms specific to particular processes in particular contexts and particular phenomenal domains. Perhaps part of the problem with the concept is that it has been approached at too abstract a level.

## The Contribution from Social Science

The micro-macro problem – the relationship between the actions of individuals and resulting social structures and the reciprocal constraint those structures place on individual agency – has long standing in social science. The problem is central to many social theories developed throughout the 19<sup>th</sup> and 20<sup>th</sup> century. Examples include: Marxian dialectical materialism [15] built upon by, among others, Vygotsky [16] and Lyont’ev [17]; the social constructionism of Berger and Luckmann [18]; Giddens’ structuration theory [19]; and the recent work of critical realists [20-23]. These alternative theories are frequently founded on differing assumptions, extending from the essentially objectivist/rationalist theory of Coleman [24], through the critical theories of Habermas and then to the radical constructivism of Luhmann [25, 26].

Fuchs & Hofkirchner [27: 33] have recently suggested a four category schema for classifying social theory according to the ontological position adopted with respect to the micro-macro relationship. The majority of existing social theories, they argue, fall into one or other of the categories: *individualism* and *sociologism*. Neither of these ‘paradigms’ provides a theoretical foundation which supports exploration let alone the possibility of advancing understanding of the interplay between agency and structure. The third category, *dualism*, while considering both aspects, maintains a dichotomous stance as necessary and again does not advance any understanding of the interplay. Only those theories categorized as *dialectical* therefore have relevance. Even here, it is reasonable to conclude that little practical advance has been achieved, as most positions result in a straddling of bottom up and top-down arguments and/or suffer from excessively vague conceptualisation. These theories quickly break down again into a dichotomy the moment an attempt is made to make them operational.

What has been largely agreed, despite the very different theoretical and often inadequate handling of this problem, is that structure and agency come together in *activity* or in *body-hood* – the specific psycho-motor state at the instant of enaction. Both Vygotsky and Giddens, for example, focus on action as the point of intersection between human agency and social structures and it is implicit in Bourdieu’s *habitus* also.

## The Contribution from Systems Theory

Systems language was evident in the work of the early Emergentists and in sociology and anthropology which took seriously the structure/agency problem – notably that of Margaret Mead and Gregory Bateson. However, ‘systems’ as a focus of research took form with Bertalanffy’s attempt to establish a General Systems Theory [28, 29]. As the science of ‘wholes’ systems theory stands in contrast to reductionisms

concern with parts: it was advanced as a counter to what was perceived as excessive reductionism dominating scientific discourse during much of the 20<sup>th</sup> century.

Early (first order) cybernetic approaches modelled systems as ‘black boxes’ effectively masking the relationship between micro and macro. Application of the concept to social science by Ernst von Glasersfeld and Heinz von Foerster [30] led to social (second order) cybernetics and soft systems approaches [31] more useful for describing the systemic behaviour of social systems. While the aspiration of the General Systems Movement to establish a general science of systems is widely regarded as having failed [32], systems approaches have contributed valuable methods for the study of the interplay between levels. The Systems view of emergence was founded on:

- Holism; the whole is greater than the sum of its parts.
- A concern with *positive and negative feedback*.
- A concern with boundaries and boundary conditions – including as an epistemic act rather than an ontological fact.

More recently the development of complex systems theory and its application to natural, social and cognitive phenomena has provided additional concepts upon which much current debate about emergence draws. Many of these concepts and methods have become widely used within the multi-agent modelling community [33-36].

In contrast to the position taken by the British Emergentists who argued that irreducibility was the *exception* [8], most real world systems are now argued to be non-linear [37-40] and hence irreducible. It is non-linearity which contributes to these system’s capacity for novelty and unpredictability through the presence of deterministic Chaos [41, 42] and/or equifinality. Equifinality refers to a system where a single high level property may be realized by more than one set of micro-states which have no lawful relationship between them [12, 43, 44]. As there is no *a-priori* basis by which the likely micro state can be determined, such systems are irreducible and unpredictable in principle.

## The Contribution of Theoretical Biology, Cognitive Science and Robotics

While complexity science has drawn on a diverse range of research threads, one area where an interest in emergent phenomena has been strongly represented is in Artificial Life [45] (Alife). While initially involving exploration of emergence using very simple ‘cellular automata’, there has been increased interest within this community to explain the fundamental building blocks of life. In contrast to first generation Artificial Intelligence [46] this has included a commitment to a bottom up methodology – i.e. evolving cognitive capability rather than engineering it in [47]. This has led the field to a biologically grounded perspective of cognition and one very different from the symbolic representation approach adopted within first generation AI. From this perspective any social emergent structures will be constrained by the biological fundamentals of cognition. In other words, behavioural and linguistic domains will depend on and be constrained by the metabolic systems which give rise to them. This has bridged Alife research into theoretical biology, in particular, autopoietic theory [47-49] and hence enactive theories of cognition [50, 51].

The enactive view of cognition was first proposed by the theoretical biologists Humberto Maturana and Francisco Varela [52, 53]. While these authors primary contribution has been towards understanding the self-organising metabolic mechanisms of life, the resulting theory of autopoietic systems provided a foundation for a general theory of cognition [54-56]. This *embodied/enactive* view stands in stark contrast to the symbolic representation [57], rational actor and game theoretical approaches which have most commonly informed social simulation. It has however recently seen considerable application in robotics [66-69], where it is argued to be fundamental to understanding how robots can become genuinely autonomous – i.e. capable of learning about their environment without the need for detailed information being provided by a designer. Within social theory some consideration has been given to the implications of enaction for understanding and theorising social behaviour [26, 58-60] although not without some controversy [61-63] and we have argued elsewhere that many of these extensions are incompatible with the original concept [64, 65]. None of this has yet found extension into social simulation.

Attempts to understand and specify mechanisms of social emergence have generally built upon the philosophical and systems theoretical literatures. There has been little accommodation of the wider debate about agency and structure particularly that associated with dialectical social theory. The micro level assumptions have been largely restricted to those associated with the rational actor and game theory and first generation AI. Very little work has been done to incorporate the perspective offered by recent developments in artificial life, robotics and theoretical biology. It is however this detailed work on the relationship between cognitive capability and associated emergent behaviour that arguably provides the most valuable contribution to our understanding of social emergence. This is in part due to it being grounded in the study of real biological entities and/or the practical challenges of building viable robots.

## 3 ORDERS OF EMERGENCE

A number of authors have identified what they refer to as orders of emergence. Gilbert, for example distinguishes between a first and second order. First order emergence includes macro structures which arise from local interactions between agents of limited cognitive range (particles, fluids, reflex action). By contrast, second order emergence is argued to arise ‘*where agents recognise emergent phenomena, such as societies, clubs, formal organizations, institutions, localities and so on where the fact that you are a member or a non-member, changes the rules of interaction between you and other agents.*’ [70]. This reflects high order cognition – in particular a capacity to distinguish class characteristics, assess ‘self’ for conformity and to change behaviour accordingly. First and second order emergence then each imply qualitatively distinct cognitive mechanisms and suggest a continuum of orders of emergence linked to cognitive capability.

In a similar vein, Castelfranchi [2: 27] has distinguished ‘cognitive emergence’ which: ‘*... occurs where agents become aware, through a given ‘conceptualization’ of a certain ‘objective’ pre-cognitive (unknown and non deliberated) phenomenon that is influencing their results and outcomes, and then, indirectly, their actions.*’ This approach is based on a first generation AI [46] approach to conceptualizing agents – agent

cognition is assumed to involve acting on beliefs desires and intentions (BDI). Thus Castelfranchi conceives of a feedback path from macro pattern to micro behaviour in much the same way as Gilbert, except that here a cognitive mechanism is specified. Castelfranchi argues that this mechanism 'characterises the theory of social dynamics' and gives rise to a distinct class of emergent phenomena. In this account, the representations agents have about the beliefs, desires and intentions of other agents plays a causal role in their subsequent behaviour and therefore shapes the structures they participate in generating. Castelfranchi argues that understanding this process is fundamental to social simulation: it is where social simulation can make its greatest contribution.

These ideas are more comprehensively reflected in the five orders of emergence suggested by Ellis [71:99-101].

Order	Ellis' Description of Properties	Characteristic Organization <sup>3</sup>
1	Bottom up leading to higher level generic properties (examples include the properties of gases, liquids and solids)	Property emergence
2	Bottom up action plus boundary conditions lead to higher level structures (e.g. convection cells, sand piles, cellular automata)	Self-organization Far-from-Equilibrium (weak-autonomy)
3	Bottom up action leading to feedback and control at various levels leading to meaningful top down action - teleonomy (e.g. living cells, multi-cellular organisms with 'instinctive' - phylogenetically determined reactive capability)	Self-production (autopoiesis) of metabolism (strong autonomy)
4	as per 3 but with the addition of explicit goals related to memory influence by specific events in the individual history (i.e. capable of learning)	Autonomous sensori-motor loops (strong autonomy)
5	In addition to 4 some goals are explicitly expressed in language (humans).	Semiotic autonomy (strong autonomy)

**Table one: Adapted from Ellis.**

In table one we set out Ellis' order number in column one and his description of the associated characteristics in column two. In column three we suggest an alternative classification which draws on the distinctions suggested by Rocha.

As with the approach of Gilbert and Castelfranchi, Ellis's framework also suggests that the range and type of emergence possible depends fundamentally on the range and class of behaviour agents are able to generate.

Considering category one emergence: particles have fixed properties and are able to enter into a limited range of interactions with others based on those properties. Swarms of particles can nevertheless demonstrate some rudimentary self-organisation and hence emergence [45]. Physics has furnished good accounts of many specific examples [73] but they have limited implication for our understanding of social behaviour.

Category two has also been well explored – it is the focus of a great deal of the work undertaken on complex, far from equilibrium systems [74, 75]. Examples include the work of Per

Bak [76] on sand piles and earthquakes and Lorenz [42] on weather systems. Many so called social simulations which incorporate agents which have fixed behaviours and no capacity for learning also arguably belong here. These include classic simulations such as Schellings segregation model, the cooperation models of Axelrod [77] or the Sugarscape models of Epstein and Axtell [78]. Some may argue that these models involve agents with goals and therefore represent examples of fourth order emergence. The transition between third order and fourth, as will be argued below, involves a move to agent autonomy that is missing in these models – their goals are designed in and not a result of their own operation – it is for this reason that we argue they belong to order two although some may argue they represent reasonable analogues of the type of behaviour that might be generated by agents with higher order capability.

It is significant that Ellis' provides primarily biological examples for his category three order of emergence. The paradigmatic biological entity which illustrates the processes of reciprocal micro-macro causality pointed to by Ellis and for which we have an excellent description which has been made operational in vitro and in silico [see for example 79, 80] is the cell. While the mechanisms of autocatalysis and the metabolic pathways of cell self-production are well documented and closely studied, the most concise articulation of the fundamental self-producing processes involved comes in the theory of autopoiesis already mentioned [52, 53, 80, 81]. Varela [82: 78] states: *Autopoiesis is a prime example of a ...dialectics between the local component levels and the global whole, linked together in reciprocal relation through the requirement of constitution of an entity that self-separates from its background.* In other words the distinguishing characteristic in this order is that the micro-macro interplay leads to an autonomous structure which acts so as to maintain its viability as an entity. This is not the case for many far from equilibrium systems such as weather systems. The maintenance of viability is a clear threshold and one we appear far from being able to simulate using existing methods.

In his third order category Ellis includes a range of capabilities of biological entities up to and including 'instinctive' action. These suggest that single and multi-cellular organisms including those with a central nervous system would all be included. It may be that this order is too broadly cast. Ellis has grouped entities such as cells which rely exclusively on metabolic self-regulation with entities which also have a capacity to self-regulate using sensorimotor mechanisms. Differentiated aggregates of cells display greater capacity to respond to their environment, even where they do not possess a central nervous system, than do individual cells (e.g. by development of an immune response). A central nervous system provides the entity with even greater behavioural plasticity [52] and hence a capacity to maintain its viability in a wider range of environments. As a consequence each threshold probably originates a distinct macro phenomenology different from that of the cells that constitute them [53].

The primary point of distinction between order three and order four would appear to be between (phylogenetically) fixed individual characteristics and a capacity for an individual agent to learn. This category covers animals up to human but this again is a big span covering a number of cognitive and developmental thresholds, including the emergence of pre-linguistic theory of mind, and self-awareness [83] which might be expected to have

<sup>3</sup> Classification according to the work of [72]. Rocha, L.M., *Language Theory: Consensual Selection of Dynamics*. Cybernetics and Systems, 1996. 27(6): p. 541-553.]

a significant effect on social emergence. It is also not clear what is meant by learning. Learning can span a wide range of capabilities from simple operant conditioning to advanced reasoning.

The final transition between order four and five demarcates the line between non-human animals and humans. The advent of language gives rise to a distinct phenomenal domain with significant implications for social emergence. This is not least due to the association language has in humans with other cognitive capabilities such as theory of mind, narrative ability and reflexivity.

Examining the characteristic organization implied in Ellis' orders of emergence shows that the transition points are strongly linked to processes of self-organisation and autonomous closure. Furthermore this autonomous closure occurs recursively: closure at one level makes possible closure at a higher level and so on. What we are essentially attempting to do in social simulation at present is to shortcut this process: to achieve reasonable analogues of behaviour at various levels without also modelling the processes upon which it depends. This appears reasonable – we do not need to model sub-atomic processes in order to work with models of molecules and understand the reaction chains they can participate in, so why would we need to model metabolic or sensorimotor systems in order to understand social interactions? How then do we advance our understanding of the effect of different cognitive capability on orders of emergence and if and when they matter?

#### 4 AGENT AUTONOMY

Robots are generally intended to be able to perform useful functions in real and complex environments. To do so they need to have a level of autonomy: a capacity to map their worlds and to decide what is important and change their behaviour accordingly. This proved computationally difficult (if not impossible) to achieve using conventional AI approaches. A breakthrough was achieved with Brooks demonstration of the power of situated cognition [84]. It is therefore no surprise that our understanding of the implications and opportunities presented by understanding cognitive autonomy has been led by the field of robotics. What then is the state of the art and what implications may it have for understanding and simulating social emergence?

In her introductory paper for the Modelling Autonomy Workshop held in San Sebastián in March 2007 (<http://www.ehu.es/ias-research/autonomy/>), Margaret Boden stated that *'very broadly speaking, autonomy is self-determination: the ability to do what one does independently, without being forced so to do by some outside power.'* She notes that the concept is problematic as there are various types and degrees of independence. This has already been illustrated above when examining Ellis' orders of cognition. In social simulation we have achieved limited independence in the form of self-organization. For Barandiaran & Moreno [85: 179], *'The main difference between self-organization and autonomy is that while self-organization appears when the (microscopic) activity of a system generates at least a single (macroscopic) constraint, autonomy implies an open process of self-determination where an increasing number of constraints are self-generated.'* This reemphasises that autonomy involves recursion: cyclic generation proceeding from simple self-organisation to closure

in a succession of phenomenal domains culminating in closure at the semiotic level.

Within Alife and robotics, it has been increasingly argued that while autopoiesis specifies the metabolic closure and self-production characteristic of living entities, cognition implies more than this. A cognitive agent has a primary autonomous metabolic loop which serves to maintain its biological viability and (at least) one other loop which links sensory surfaces with motor surfaces [see also 49]. This second loop affords the agent significant additional plasticity. This plasticity is realised within a behavioural rather than a metabolic phenomenal domain [86: 168]. The two are interdependent in that the range of behaviour the agent can generate is dependent on its biology, while its biological viability can depend on the behaviour: the recognition and escape from threat or the location of food for example. While Duijn et al argue that this sensorimotor loop is already present in the two component signal transduction system (TCST) system found in bacteria, Moreno et al [47] argue that it is the central nervous system which fundamentally distinguishes biological/metabolic processes from cognitive processes.

Irrespective of where this line is drawn, both are consistent in the view that *'...cognition is not so much a centralized property of the biological hardware of an organism, or a set of internally computed algorithms, but instead denotes an abstraction of organism environment reciprocity'*. [87]. This is consistent with the position taken by Varela [50, 82, 88], that autonomous agents *'bring forth a world'* as a result of their operational closure. In other words, what an agent can perceive and cognize is determined by its own operation, not the environment. Again from Barandiaran, under conditions of autonomy: *'It is not the organism that matches the environment in a given specified way. On the contrary it is through the particular way in which the agent satisfies the homeostatic maintenance of essential variables that an adaptive environment (a world) is specified - cut out from a background of unspecific physical surroundings.'* [49]

What this means is that the environment is a source of perturbations which act only as triggers for change. It is the nervous system's structure that dictates which perturbations can be a trigger [57, 89]. Consequently changes to the structure of one agent's nervous system, and consequently its behaviour, will be unique to that agent. The environmental perturbations that act as a change trigger in one agent will not necessarily trigger a change in another, or if they do, the change that is triggered may take a different form and/or have different implications for the viability of that agent in its environment, given its history of interactions.

The consequence of this recursive construction of increasing order of autonomy for the agent is enhanced viability in a wider range of environments. This is apparent if we consider the effect of the transition from metabolic autonomy (autopoiesis) to sensorimotor autonomy supported by a central nervous system. The coexistence of these two interdependent levels of autonomous functioning allows the organism to exploit the rapid response times of the neural system and this makes possible a significantly increased set of possible responses to environmental perturbations [49]. An organism that relies less on the slow diffusion reactions associated with metabolism, and which can draw on the rapid response of the chemical/electrical nervous system is better able to survive in less stable

environments. In systems terms, it has greater requisite variety [90].

It is this asymmetry between the state space of possible configurations made possible by an advanced nervous system and the range of response needed to maintain immediate regulation that gives rise to what we call 'agency'. Agency is a consequence of autonomy. Agency makes possible what we typically regard as distinguishing features of social systems: endogenous goal making and seeking behaviour. Agency also supports 'free-will', the opportunity for agents to behave in ways which are non-deterministic: to generate new bottom up solutions to situations they encounter. From this perspective then autonomy is fundamental to agency and hence to a capacity to engage in activity which can genuinely be called social. We conclude therefore that to be able to simulate an agent which is deserving of the title of being a social agent, it would need to exhibit some level of 'strong autonomy' and hence agency. But we are a long way from achieving it.

Why is this necessary and what does it reveal about the fundamental mechanisms at work in social systems? Also, at this stage we have considered only agents in isolation. What happens when we bring multiple autonomous agents together such that they can interact?

## 5 MECHANISMS OF SOCIALITY

Following the line of argument developed above, when brought together each agent treats each other agent as a part of its environment.

As agents interact each undergoes a set of internal structural accommodations which allow it to persist in its relationship with the others. This results in a 'structural drift', or a gradual change to the state of each agents nervous system [52, 53]. Over time it traces a unique history – Maturana refers to this as the agent's *ontogeny*. When interactions become 'recurrent' – that is repetitive and ongoing – agents can become 'structurally coupled'. Here we have the most basic element of sociality and one that can be applied to all organisms with nervous systems, even very elementary ones.

Importantly, a history of recurrent interactions leads to a structural congruence or commonality of experience between two or more agents: their behaviours become tuned to one another in a reciprocal 'dance' maintained in and through their relating. The degree of structural coupling that arises when two or more agents interact is a fundamental factor in determining the dynamics and emergent behaviour of the resulting structurally coupled system. Agents give rise to a behavioural phenomenal domain in which a range of attractors may form. These attractors are what we would typically call macro-social structures. What then of the advent of language?

Language is associated with higher order cognition. It will support a behavioural domain which is more inherently plastic than one coordinated only through bodily interaction. Otherwise, as a mechanism, it is an extension of what has already been discussed. It is however a non trivial extension as the state space of points of interaction becomes very much larger where the variables (utterances) are recursive as they are in language as a) agents make linguistic distinctions on linguistic distinctions b) new linguistic constructs are under control of the system they also serve to regulate.

Structural coupling within a linguistic domain will be apparent from the convergence of the individual linguistic utterances to form a shared lexicon and grammar. The driving force behind this convergence is the one fixed internal goal the agents have: that of maintaining their viability. If they are biological agents this will involve the preservation of their autopoiesis – i.e. remaining alive. At base level this involves meeting the requirements of the metabolic level of operation. They will need to eat, stay warm, avoid predators and find partners. The metabolic or biological necessarily interacts with the behavioural and the linguistic domains: the domains are co-dependent.

The agents will innovate in their behaviour in order to satisfy their minimal requirements. Some of the behaviours they adopt will, however, be due to a need to accommodate the behaviours of other agents. A set of attractors should therefore emerge which represent sets of states which 'satisfice' social constraints as well as fundamental biological constraints. To an observer<sup>4</sup> some of these states may appear as goal based (food seeking) while others may be seen to be primarily to do with mutual accommodation (norms). The attractors may be reflected in macro structuration (division of labour, identity groups) and may assemble into yet higher order patterns (organizations, institutions). The engine of this process of social emergence is structural coupling and the dimensions of possible coupling and the scope of behaviours which may be involved in establishing and maintaining coupling is dependent on the biology and cognitive plasticity of the agent.

We are currently able to simulate behaviours up to order three. Moving beyond this raises some interesting questions. Among these are: are the cognitive capabilities clearly associated with social behaviour necessarily tied to metabolic autonomy? To what degree do these capabilities manifest the way they do due to the specific organic mechanisms associated with life? Is it possible to simulate behavioural autonomy and linguistic systems which are operationally closed on other than an organic substrate? If so what are the essential low level characteristics which are essential to supporting them? In short: is it possible to model these types of processes in-silico?

## 6 CONCLUSION AND FUTURE DIRECTIONS

There is a range of ways of thinking about the relationship between micro and macro level phenomena. There have been centuries of debate about the relative merit of reductionist, vitalist and holistic perspectives for understanding how higher order structures emerge from lower. Despite ongoing scepticism in some philosophical quarters, we have advanced our understanding of the mechanisms involved to a very significant degree over the past 30 years. Emergent structures are increasingly understood to be a product of non-linear interactions associated with complex systems of agents. We can however, go further than this. One of the insights being

---

<sup>4</sup> The changing role of the observer in distinguishing different categories of emergence is a significant issue which has not been taken up here. Arguably observer based distinctions play no role in the property emergence associated with Ellis' category one, whilst they are intrinsic to the social emergence associated with category five. The role in the intermediate categories is less clear. We have set out some early thinking about the critical role this plays in 91. Goldspink, C. and R. Kay. *Social Emergence: Distinguishing Reflexive and Non-reflexive Modes in AAAI Fall Symposium 2007*. Washington.

developed is that the range and type of emergent structure depends on the specific mechanisms involved and on the properties of the micro agents.

This paper has concentrated on how alternative micro-capabilities support qualitatively distinct forms of social emergence. What has been argued is that social emergence implies not a single transition from micro to macro but is built upon, and is an example of recursive self-organization within a biological domain. The recursive levels in living systems span metabolic, neurological, social-behavioural and social-semiotic levels.

Social emergence involves a level of self-reference and self-generation which is not apparent in non-organic forms of far from equilibrium behaviour. Social emergence builds on biological emergence. This is to say that the phenomenal domains associated with social systems, particularly those involving humans, are constrained by the biological processes which make them possible. The ongoing debate about emergence as a concept demonstrates that understanding the relationship between micro and macro phenomena is theoretically as well as practically challenging. To date it has proven difficult to build models which provide reasonable analogues of this process. What has been achieved has been achieved largely in robotics, and Artificial Life. So far, social simulation has played a minor role. Nevertheless the science of these processes is important to social simulation. It has proven possible to model some social behaviour to good effect without agents with these capabilities. It can also be argued that in highly complex (chaotic or random) environments higher order cognition is of little value justifying a parsimonious substitution of particle-like agents. However, it is reasonable to expect that we will not be able to effectively model some forms of social behaviour without having come to terms with and found ways to simulate behaviour which is possible due to autonomous closure. Equally social simulation could play an important role in helping us to understand the implications of autonomous closure and for advancing our ability to theorise about it.

## References

1. Sawyer, K.R., *Artificial Societies: Multiagent Systems and the Micro-macro Link in Sociological Theory*. Sociological Methods & Research, 2003. **31**: p. 38.
2. Castelfranchi, C., *Simulating with Cognitive Agents: The Importance of cognitive emergence*, in *Multi-agent Systems and Agent Based Simulation*, J.S. Sichman, R. Conte, and N. Gilbert, Editors. 1998, Springer: Berlin.
3. Goldspink, C. and R. Kay, *Organizations as Self Organizing and Sustaining Systems: A Complex and Autopoietic Systems Perspective*. International Journal General Systems, 2003. **32**(5): p. 459-474.
4. Goldspink, C. and R. Kay, *Bridging the Micro-Macro Divide: a new basis for social science*. Human Relations, 2004. **57** (5): p. 597-618.
5. Ablowitz, R., *The Theory of Emergence*. Philosophy of Science, 1939. **6**(1): p. 16.
6. Peterson, G.R., *Species of Emergence*. Zygon, 2006. **41**(3): p. 22.
7. Shrader, W.E., *The Metaphysics of Ontological Emergence*, in *Graduate School*. 2005, University of Notre Dame.
8. Eronen, M., *Emergence in the Philosophy of Mind*, in *Department of Philosophy*. 2004, University of Helsinki: Helsinki. p. 79.
9. Stanford Encyclopaedia of Philosophy, *Emergent Properties*, in *Stanford Encyclopaedia of Philosophy*. 2006.
10. Clayton, P. and P. Davies, *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. 2006, Oxford: Oxford University Press.
11. Fodor, J.A., *Special; Sciences or The Disunity of Science as a Working Hypothesis*. Synthese, 1974. **28**: p. 18.
12. Sawyer, K.R., *Emergence in Sociology: Contemporary Philosophy of Mind and Some Implications for Sociology Theory*. American Journal of Sociology, 2001. **107**(3): p. 551-585.
13. Symons, J., *Emergence and Reflexive Downward Causation*. Principia, 2002. **1**: p. 183-202.
14. Davies, P., *The Physics of Downward Causation*, in *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*, P. Clayton and P. Davies, Editors. 2006, Oxford University Press: Oxford.
15. Engels, F., *Dialectics of Nature*. 1934, Moscow: Progress Publishers.
16. Vygotsky, L.S., *Thought and Language*, ed. T.b.E.H.a.G. Vakar. 1962, Cambridge, Mass: MIT Press.
17. Leont'ev, A.N., *Activity, Consciousness and Personality*. 1978, Engelwood Cliffs: Prentice Hall.
18. Berger, P.L. and T. Luckman, *The Social Construction of Reality*. 1972: Penguin.
19. Giddens, A., *The Constitution of society: Outline of the theory of structuration*. 1984, Berkeley: University of California Press.
20. Bhaskar, R., *A Realist Theory of Science*. 1997, Verso: London.
21. Bhaskar, R., *The Possibility of Naturalism*. 1998, London: Routledge.
22. Archer, M., *Realism in the Social Sciences*, in *Critical Realism: Essential Readings*, M. Archer, et al., Editors. 1998, Routledge: London.
23. Archer, M., et al., *Critical Realism: Essential Readings*. 1998, London: Routledge.
24. Coleman, J.S., *Foundations of Social Theory*. 1994, Cambridge: Belknap.
25. Luhmann, N., *Essays on Self Reference*. 1990, New York: Columbia University Press.
26. Luhmann, N., *Social Systems*. 1995, Stanford: Stanford University Press.
27. Fuchs, C. and W. Hofkirchner, *The Dialectic of Bottom-up and Top-down Emergence in Social Systems*. tripleC 2005. **1**(1): p. 22.
28. Bertalanffy, L.v., *An Outline of General Systems Theory*. British Journal for the Philosophy of Science, 1950. **1**(2).
29. Bertalanffy\_von, L., *General Systems Theory*. 1968, New York: Braziller.
30. Keeney, B.P., *Aesthetics of Change*. 1987: Guilford.
31. Checkland, P., *Systems Thinking Systems Practice*. 1988, G.B.: John Wiley.
32. Jackson, M.C., *Systems Approaches to Management*. 2000, London: Kluwer Academic.
33. Holland, J.H., *Emergence: from chaos to order*. 1998, Ma.: Addison Wesley.

34. Gilbert, N., *Emergence in Social Simulation*, in *Artificial Societies*, N. Gilbert and R. Conte, Editors. 1995, UCL Press: London.
35. Castelfranchi, C., *Simulating with Cognitive Agents: The Importance of Cognitive Emergence*, in *Lecture Notes in Artificial Intelligence*, J.S. Sichman, R. Conte, and N. Gilbert, Editors. 1998, Springer Verlag: Berlin.
36. Conte, R., R. Hegselmann, and P. Terna, *Simulating Social Phenomena*. 1997, Berlin: Springer.
37. Stewart, I., *Does God Play Dice - The New Mathematics of Chaos*. 1990: Penguin.
38. Kauffman, S., *Investigations*. 2000, New York: Oxford.
39. Kauffman, S.A., *The Origins of Order: Self Organization and Selection in Evolution*. 1993: Oxford University Press.
40. Kauffman, S.A., *At home in the Universe: The Search for Laws of Complexity*. 1996, London: Penguin.
41. Williams, G.P., *Chaos Theory Tamed*. 1997, Washington D.C: Joseph Henry Press.
42. Lorenz, E.N., *The Essence of Chaos*. 4 ed. 2001, Seattle: University of Washington Press.
43. Richardson, K.A., *Methodological Implications of a Complex Systems Approach to Sociality: Some further remarks*. *Journal of Artificial Societies and Social Simulation*, 2002. **5**(2).
44. Richardson, K.A. *On the Limits of Bottom Up Computer Simulation: Towards a Non-linear Modeling Culture*. in *36th Hawaii International Conference on Systems Science*. 2002. Hawaii: IEEE.
45. Kennedy, J. and R.C. Eberhart, *Swarm Intelligence*. 1 ed. 2001, London: Academic Press. 511.
46. Franklin, S., *Artificial Minds*. 1998, London: MIT press.
47. Moreno, A., J. Umerez, and J. Ibanes, *Cognition and Life*. *Brain and Cognition*, 1997. **34**: p. 107-129.
48. Beer, R.D., *Autopoiesis and Cognition and the Game of Life*. *Artificial Life*, 2004. **10**: p. 309-326.
49. Barandiaran, X., *Behavioral Adaptive Autonomy. A Milestone on the ALife route to AI?* 2005, Department of Logic and Philosophy of Science, University of the Basque Country: San-sebastian, Spain.
50. Rudrauf, D., et al., *From Autopoiesis to Neurophenomenology: Francisco Varela's exploration of the biophysics of being*. *Biol. Res*, 2003. **36**: p. 27-65.
51. Thompson, E. and F.J. Varela, *Radical embodiment: neural dynamics and consciousness*. *TRENDS in Cognitive Sciences* 2001. **5**(10): p. 7.
52. Maturana, H. and F. Varela, *Autopoiesis and Cognition: The Realization of the Living*. *Boston Studies in the Philosophy of Science*. Vol. 42. 1980, Boston: D. Reidel.
53. Maturana, H.R. and F.J. Varela, *The Tree of Knowledge: The Biological Roots of Human Understanding*. 1992, Boston: Shambhala.
54. Maturana, H., *Ontology of Observing: The Biological Foundations of Self-Consciousness and of The Physical Domain of Existence*. 1988.
55. Maturana, H., *The origin and conservation of self-consciousness Reflections on four questions by Heinz von Foerster*. *Kybernetes* 2005 **34**(1/2): p. 34.
56. Maturana, H., J. Mpodozis, and J.C. Letelier, *Brain, Language and the Origin of Human Mental Functions*. *Biological Research* 1995. **28**: p. 11.
57. Varela, F., E. Thompson, and E. Rosch, *The Embodied Mind*. 1992, Cambridge: MIT Press.
58. Teubner, G., *How the Law Thinks: Towards a Constructivist Epistemology of Law*. *Law and Society Review*, 1989. **23**(5): p. 727-758.
59. Zeleny, M., *Autopoiesis: A Theory of Living Organization*. 1991, New York: North Holland.
60. von Krogh, G. and J. Roos, *Organizational Epistemology*. 1995, London: St Martins Press.
61. Mingers, J., *Are Social Systems Autopoietic? Assessing Luhmanns Social Theory*. *Sociological review*, 2002. **50**(2).
62. Mingers, J., *Can Social Systems be Autopoietic? Bhaskar's and Giddens' Social Theories*. *Journal for the Theory of Social Behaviour*, 2004. **34**(4): p. 25.
63. Bednarz, J., *Autopoiesis: the Organizational Closure of Social Systems*. *Systems Research*, 1988. **5**(1): p. 57-64.
64. Goldspink, C., *Social Attractors: An Examination of the Applicability of Complexity theory to Social and Organisational Analysis*, in *Social Ecology*. 2000, University Western Sydney: Richmond. p. 312.
65. Kay, R., *Towards an autopoietic perspective on knowledge and organisation*, in *Social Ecology*. 1999, University of Western Sydney: Richmond.
66. Di Paolo, E.A., *Autopoiesis, adaptivity, teleology, agency*. n.d, Department of Informatics, University of Sussex: Brighton.
67. Di Paolo, E.A., *Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop*. n.d, School of Cognitive and Computing Sciences, University of Sussex: Brighton.
68. Di Paolo, E.A. and H. Lizuka, *How (not) to Model Autonomous Behaviour*. *Biosystems*, 2007.
69. Di Paolo, E.A., M. Rohde, and H. De Jaegher, *Horizons for The Enactive Mind: Values, Social Interaction and Play*, in *Enaction: Towards a New Paradigm for Cognitive Science*, J. Stewart, O. Gapenne, and E.A. Di Paolo, Editors. 2007, MIT Press: Cambridge MA.
70. Gilbert, N. *Varieties of Emergence*. in *Social Agents: Ecology, Exchange, and Evolution Conference 2002*. Chicago.
71. Ellis, G.F.R., *On the Nature of Emergent Reality*, in *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*, P. Clayton and P. Davies, Editors. 2006, Oxford University Press: Oxford.
72. Rocha, L.M., *Language Theory: Consensual Selection of Dynamics*. *Cybernetics and Systems*, 1996. **27**(6): p. 541-553.
73. Gell-Mann, M., *The Quark and the Jaguar: Adventures in the simple and the complex*. 1995, Great Britain: Abacus.
74. Prigogine, I., *The End of Certainty: Time, Chaos and the New Laws of Nature*. 1997, New York: The Free Press.
75. Prigogine, I. and I. Stengers, *Order out of Chaos: Man's New Dialogue with Nature*. 1985: Flamingo.
76. Bak, P., *How Nature Works: The Science of Self-Organized criticality*. 1996, New York: Copernicus.
77. Axelrod, R., *The Evolution of Cooperation*. 1984, New York: Basic Books.
78. Epstein, J.M. and R. Axtel, *Growing Artificial Societies*. 1996, Cambridge, Ma.: MIT Press.
79. McMullin, B. and D. Grob, *Towards the Implementation of Evolving Autopoietic Artificial Agents*, in *6th European Conference on Artificial Life ECAL 2001*. 2001: University of Economics, Prague.

80. Varela, F., H. Maturana, and R. Uribe, *Autopoiesis: The Organization of Living Systems, Its Characterization and a Model*. Biosystems, 1974. **5**: p. 187-196.
81. Varela, F., *Principles of Biological Autonomy*. 1979, New York: Elsevier-North Holland.
82. Varela, F., *Patterns of Life: Intertwining Identity and Cognition*. Brain and Cognition, 1997. **34**: p. 72-87.
83. Gardenfors, P., *How Homo became Sapiens: On the evolution of Thinking*. 2006, Oxford: Oxford University Press.
84. Brooks, R.A., *Intelligence without representation*. Intelligence without reason, 1991(47): p. 569-595.
85. Barandiaran, X. and A. Moreno, *On what makes certain dynamical systems cognitive: A minimally cognitive organization program*. Adaptive Behavior, 2006. **14**(2): p. 171-185.
86. Moreno, A. and A. Etxeberria, *Agency in natural and artificial systems*. 1995, Department of Logic and Philosophy of Science University of the Basque Country: San Sabastian, Spain.
87. Duijn, M.v., F. Keijzer, and D. Franken, *Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination*. Adaptive Behavior, 2006. **14**(2): p. 157-170.
88. Thompson, E., *Life and Mind: From Autopoiesis to neurophenomenology, a Tribute to Francisco Varela*. Phenomenology and the cognitive Sciences, 2004. **3**: p. 381-398.
89. Mingers, J., *The Cognitive Theories of Maturana and Varela*. Systems Practice, 1991. **4**(4): p. 319-338.
90. Ashby, W.R., *Self-regulation and Requisite Variety*, in *Systems Thinking*, F.E. Emery, Editor. 1974, Penguin: Great Britain.
91. Goldspink, C. and R. Kay. *Social Emergence: Distinguishing Reflexive and Non-reflexive Modes in AAAI Fall Symposium 2007*. Washington.