research proposal submitted for funding to the FWO by F. Heylighen, G. Geeraerts and F. Van Overwalle

Mediated evolution of social organisation: a multi-agent simulation

INTRODUCTION

Rapid technological developments together with the process of globalisation bring about an entangled dynamic of contradictory forces: integration and differentiation, competition and co-operation, individualism and solidarity. Our society seems to be in a very unstable transitional phase, so that it is difficult to conceive how the desired result—a sustainable, integrated world system respecting individual and collective diversity (Heylighen, 2003; Stewart, 2000)—will come about.

The ever increasing complexity, changeability and unpredictability make it particularly difficult to understand and manage this social evolution. Traditionally the social sciences have focused on specific sub-problems, such as the relationships between social classes or ethnic groups, or the tension between market and solidarity. Yet the problem of complex social evolution is too encompassing to be approached through a single aspect, model or sub-problem. That is why we need a broad, interdisciplinary approach, that focuses on the abstract essence: how does social organisation arise and evolve, i.e. how do initially autonomous actors, through mutual interaction, come to form an increasingly complex, integrated system?

In essence there exist two, complementary approaches to this problem, that stress, respectively, conflict and the dynamics it creates, and cooperation and the system of mutual dependencies that it entails. The first can be found in conflict sociology, classical economics which sees all actors as competitors, and the Darwinist theory of evolution which starts from a "struggle for life" between competing individuals. The second approach lies at the basis of functionalism in sociology and of the general systems theory. This approach sets out from a system or synergetic whole, characterised by emergent properties that cannot be reduced to the properties of the parts. But the systems approach has not advanced much beyond a static description of systems on different levels of complexity, lacking an explanation of how these come into existence. The conflict and evolutionary approaches on the other hand are more dynamic, proposing models (e.g. based in game theory) of the evolution of cooperation starting from initially selfish actors (e.g. Axelrod, 1984). Still, the problem remains that this approach is in essence reductionist: all phenomena are reduced to the interactions between individual actors (methodological individualism, cf. Heylighen & Campbell, 1995). Emergent organisation is not explained.

AIM

We hold that these issues of conflict and cooperation can be resolved in the broadest sense by synthesising the evolutionary and the systems approaches. We intend to approach this synthesis through its core issue, namely the analysis and understanding of evolutionary transitions or metasystem transitions (Turchin, 1977; Maynard-Smith

& Szathmary, 1995; Heylighen, 2000; Michod, 1999), i.e. the fundamental processes through which an integrated system arises.

Examples include the origin of life, the transition from unicellular to multicellular organisms, and from individuals to societies. Common property of such transitions is that systems which initially were able to survive and reproduce autonomously, consequently have become dependent on a larger synergetic whole. A number of these wholes can in a later stage again be joined together, forming a supersystem of an even higher order. Subsequent transitions explain the fundamentally hierarchical evolution of complex systems, and indicate the general trend of increasing complexity, organisation and synergy that characterises evolution (Turchin, 1977; Stewart, 2000; Heylighen, 1999).

The current project aims at the integration of the different approaches to this issue, that have largely originated independently (Heylighen, 2000) in disciplines such as biology, cybernetics, and political science. Our focus will be on the general dynamics of the evolution of complex organisation, and on the sometimes contradictory values this entails that implicitly drive social and biological systems. Thus, our analysis should contribute to a better understanding of the very complex dynamics of integration and differentiation, competition and solidarity that characterise the present process of globalization (Stewart, 2000).

STARTING HYPOTHESES

In order to make this very broad and abstract issue more concrete, we want to base our research on a number of specific working hypotheses, building on our previous work with the famous methodologist, the late Donald T. Campbell (Heylighen & Campbell, 1995) and our most recent "mediator model" (Heylighen, 2004).

According to basic Darwinist principles, individual systems are fundamentally "selfish" (Dawkins, 1989), in the sense that they were selected to maximise their own fitness, not the fitness of the group to which they belong. In a hierarchical system, each of the levels is subject to natural selection, but these selective pressures are often contradictory, since what is "best" (fitness-maximizing) for a subsystem, is not in general "best" for the supersystem. This produces an inherent tension, which can be found e.g. in the ambivalence of human psychology, which constantly vacillates between selfishness and altruism, or competition and co-operation. The same tensions can be found at other levels of complexity, such as the cell that must "choose" between submission to the organism (e.g. by undergoing apoptosis—"programmed suicide") and the opportunity to maximally reproduce, giving rise to a tumour.

Evolutionary theorists have recently proposed different models that explain how selfish components can still manage to form a supersystem that is fit on the group level, in spite of the contradictory forces exerted on them. Yet there remains a fundamental impediment to the maintenance of such a system: the "free rider" problem (Dawkins, 1989; Heylighen & Campbell, 1995). Even though it is advantageous for group members to cooperate, the greatest benefit will still go to the free riders, who profit from the effort of others without investing anything in return. (e.g. countries that step out of a global convention, such as the Kyoto Treaty, because that costs them less). Because of their higher individual fitness, free riders will outcompete the altruists, and thus destroy the cooperative system from within–unless there exists a mechanism that limits their freedom of movement.

Such control mechanisms are found at all levels of biological and social complexity (Maynard Smith & Szathmary, 1995; Stewart, 2000), but there is as yet no universal explanation for their origin. Various partial explanations have been postulated for specific cases, including group selection, kin selection, reciprocal altruism, game theory, internalised morality, market mechanisms, norms, laws and institutions (Heylighen & Campbell, 1995).

The more general model we wish to develop starts from the interactions between initially autonomous actors or agents (cells, individuals, primary groups, firms, countries...). Since these agents to some extent require the same resources to survive and grow, their primary interaction is competition, i.e. the action of an agent to acquire more resources will in general be to the detriment of other agents. Yet such a conflictual interaction (a zero-sum game) is evolutionarily unstable, since it limits overall fitness. Natural selection prefers synergy, i.e. interactions that produce benefits for all the agents (positive-sum games). Therefore the evolution of actions will tend to produce patterns that limit conflict, and that increase mutual benefit (Wright, 2000). This can be seen as the self-organisation of interaction patterns towards a more synergetic, coordinated system. But in order to stabilize this synergy and protect it from free riders, we need a reliable steering mechanism, i.e. a control system independent of individual agents. Following the terminology of Stewart (2000), we will call this system the "manager". According to our working hypothesis (Heylighen, 2004), this system evolves in three consecutive stages.

In a first stage, the interaction pattern functions as a medium enabling synergetic interaction. Examples of such media are natural language, the market, and electronic media such as the internet. In a second stage, the medium evolves into a mediator, which not only enhances synergy, but inhibits conflict. Examples of such mediators are moral rules and conventions expressed in language, money as a conventional means of value determination in a market, and standards for the use of electronic media. In the final stage, the mediator evolves into a manager, which not only coordinates spontaneous interactions, but initiates new actions when these are to the benefit of the agent collective. Examples are institutions that monitor the implementation of rules, the "invisible hand" of the market that makes supply match demand, and the "global brain", the intelligent system that according to some futurologists will emerge from the Internet. The manager turns the group of agents into an integrated, organised system, that pursues the collective interest.

Although we can find numerous specific examples of this type of self-organisation of a group of agents, the current project intends to analyse and model this process at the most general, abstract level. This will allow us to subsequently model fundamental cases, such as the evolution of language, institutions, culture, collective intelligence in insects, markets…, as specific instances of this general scenario.

METHODOLOGY

In addition to a theoretical analysis based on a literature review, and a thorough analysis of specific cases, these hypotheses will primarily be developed with the aid of computer simulations. Simulation as a tool for the analysis of social systems has become very easy and popular in the last few years, as illustrated by the numerous publications in the Journal of Artificial Societies and Social Simulation.

We plan to start from the KEBA system developed by C. Gershenson (2002) of the Center Leo Apostel, a virtual environment where agents interact with each other and with external objects, while their actions can be "rewarded" (reinforced) or "punished" (weakened) depending on the fitness benefits they bring about. By adjusting the rules that guide agent interaction, an evolutionary dynamics (Michod, 1999) can be created that promotes metasystem transitions, in which the agents come to cooperate and form a part of an integrated system. In that way different variants of the hypotheses can be implemented and visualised, so that their respective advantages and disadvantages can be explicitly observed. In particular we want to implement two concrete models for the emergence of social organisation: 1) a spontaneous differentiation into "ethnic" or "social" subgroups; 2) a coordinated division of labour, characterised by an efficient "workflow".

For the first model we start from interactions in which an agent initiates an interaction and the other agent either complies with it (positive, cooperation) or rejects it (negative, conflict). In addition a number of "tags" or "markers" are accorded randomly to the agents, allowing them to recognise other agents. Agents learn from their interactions in the following manner: if the result is positive, the agent will get more "trust" in, or "sympathy" for, the other agent; thus the probability increases that it will react positively to actions of that agent in the future, or initiate a new interaction itself. Vice-versa, a negative result will lead to more "distrust" and a reduced probability of interaction.

But to recognise this agent, it has to take its clue from the tag, which in general is not uniquely distinguishable. This means that a later interaction may well be initiated with a different agent that carries the same (or a similar) tag, but that is not necessarily willing to cooperate to the same extent. We expect that if the first few interactions with agents with similar tags all generate positive (negative) results, the agent will develop a strong "prejudice" to always react positively (negatively) to agents characterised by that marker. Because of the positive (negative) reactions of the first agent, agents with that marker will also tend to react more positively (negatively) to this agent, and others with a similar marker.

Our expectation is that in this way, through positive feedback, the initially random interactions will produce a differentiation in clusters of similarly marked agents that cooperate with each other, but that are reluctant to interact with members of other groups. The tags and their associations thus develop the function of a self-organising mediator that increases the probability of positive interactions by creating a division between "friends" (in-group) and "strangers" or "foes" (out-group). An interesting research question is in how far this dynamics will also produce “outcasts”, who cooperate with nobody, “hubs”, who have many friends all around, or “intermediaries” who link two groups.

The second model is more ambitious, in the sense that we wish to evolve a mediator that gives the group as a whole a form of "collective intelligence" or "distributed cognition", i.e. a form of organisation that allows the agents to collectively solve problems that are too complex to be tackled individually. These problems are modelled as a task complex. The tasks are mutually dependent in the sense that a certain task or certain tasks (e.g. building a house) have to be accomplished before another task (e.g. installing electricity) can be initiated. Each agent can either execute a task itself, or delegate (forward) it to another agent.

Initially all agents are equally competent or incompetent, meaning that they have the same probability of succesfully accomplishing a task. However, each time it accomplishes a task, an agent becomes more "experienced" so that the probability increases that it will bring the same task to a successful end later on. Simultaneously the agent who delegated the task will get more trust in the competence of the delegate, and thus increase its probability to delegate a similar task to the same agent in the future. As demonstrated by the simulation of Gaines (1994), these assumptions are sufficient to have a division of labour evolve spontaneously between agents, resulting in a social differentiation analogous to that of model (1).

However, when the tasks are mutually dependent, it is not sufficient to select the right "expert" to carry out a task: first the preparatory tasks have to be performed by the right specialists, in the right order. Since the agents do not know a priori what the right order is, they can only randomly attempt to execute or delegate a task, or, failing these, pick out another task. Sooner or later they will find a task they can execute, either because it requires no preparation, or because a preparatory task has already been performed by another agent. Each accomplished task enables the accomplishment of a series of directly dependent tasks, and in this way the overall problem will eventually be solved. In each problem cycle agents will learn better when to take on which task by themselves, or when to delegate it to a specific other agent.

We expect that this learned organisation will eventually stabilise into a system of efficient, coordinated actions and division of labour, adapted to the task structure. Although no single individual agent knows the entire task structure, it can be said that the evolved knowledge is "distributed" throughout the collective. The "tags" that identify agents and the learned associations between a tag and the competence for a particular task again play the role of the interaction medium, which in this case also evolves the functions of mediator and "manager", that is to say it initiates the actions of the agents and it steers them so that the problem is tackled as efficiently as possible.

For both models our research will consist in registering and analysing the dynamics of the process of social organisation as accurately as possible, by comparing the structures that emerge during the different stages of the process. In addition, different variations of the model will be tested, inspired by alternative theoretical hypotheses coming from the literature or from our own research, and by the results of preceding simulations. Specific properties that will be varied are the numbers of agents, forms of interaction (cooperation, indifference and/or conflict), strength and dynamics of trust relationships, task structure (complexity, mutual dependency), and tag distributions (fixed or variable, random or dependent on previous interactions, more or less homogeneous). This will allow us to better understand which factors contribute to an efficient organisation, and which will rather increase the risk of conflicts, fragmentation, or prejudice.

PROJECT STAGES

2005: comparative literature review of relevant models in various disciplines; detailed specification of the basic principles of the model
2006: concrete implementation, programming, and testing of the computer simulation
2007: collection and processing of the simulation data for the different variations of the model
2008: synthesis and interpretation of the results; publication through a doctoral dissertation and papers

COORDINATION BETWEEN THE PARTICIPANTS

The project will be coordinated at the Center Leo Apostel, led by F. Heylighen and supported by C. Gershenson, with a focus on the evolutionary dynamics and computer simulation of the model. The Center for Polemology, led by G. Geeraerts and supported by K. Laforce, will analyse the dynamics of conflict and cooperation, and compare the developing model to existing political science models. The Lab for Social Psychology, led by F. Van Overwalle and supported by M. Heath, will focus on group dynamics and distributed cognition, compare the model to psychological models in this area, and participate in the design of the computer simulation. All participants in the project will remain continually in direct communication through group discussions, seminars, and e-mail.

REFERENCES

Axelrod, R. M. (1984) *The Evolution of Cooperation*. Basic Books, New York.

Gershenson C., (2002), “Behaviour-based Knowledge Systems”, Proc. 2nd Int. Workshop on Epigenetic Robotics. Edinburgh.

Gaines, B. R. (1994) “The Collective Stance in Modeling Expertise in Individuals and Organizations”, Int. J. Expert Systems 71, 22-51.

Heylighen F. & Campbell D.T. (1995): “Selection of Organization at the Social Level”, World Futures 45, p. 181-212.

Heylighen F. (2000): “Evolutionary Transitions: how do levels of complexity emerge?”, Complexity 6 (1), p. 53-57

Heylighen F. (2003): “The Global Superorganism”, Journal of Social and Evolutionary Systems [in press]

Heylighen F. (2004): “Mediator Evolution”, Artificial Life [submitted]

Maynard Smith J. & Szathmáry E. (1995): “The Major Transitions in Evolution”, W.H. Freeman, Oxford

Michod, R. E. (1999). “Darwinian Dynamics, Evolutionary Transitions in Fitness and Individuality”, Princeton University Press

Stewart, J. (2000): “Evolution's Arrow: The direction of evolution and the future of humanity”, Chapman Press, Canberra

Turchin, Valentin: The Phenomenon of Science. A cybernetic approach to human evolution 1977 (Columbia University Press, New York).

Wright, R. (2000): “Non-Zero. The Logic of Human Destiny”, Pantheon Books