Social Ethic Behavior Simulation Project

Krzysztof Cetnarowicz, Gabriel Rojek, Jan Werszowiec Płazowski, Marek Suwara, Cracow

1. Is ethics a part of personal knowledge and can there be an objective knowledge of ethical rules?

Ethics has usually been considered as the domain of the intrinsic personal belief. Some even claimed that no objective knowledge of ethics is possible [1]. We propose a quite new way of approaching the problem. Although ethics as a part of the personal belief cannot be examined scientifically, the claim that it is not possible to study ethical rules as means of strategy choice is false. The model we bring forward handles the role of ethical rules from the perspective of evolutionary fitness [2].

Individuals act within a certain society, which is influenced rather by cultural "memes" than by individually gained experience. Psychologists would claim that an individual is a necessary source of any changes in culture, because of his/her creative capabilities. From evolutionary perspective such creations are, however, merely mutations in the memotype of the culture. These mutations are subject to selection processes, which do not depend on personal characteristics of individuals but rather on the fitness to the widely understood environment. This environment includes the material world, for which fitness of a cultural "meme" means "empirical truth", as well as the cultural environment, for which the term fitness of a cultural "meme" refers to logical coherence with other "memes".

Evolutionary approach provides a researcher with several advantages:

- it is possible to abstract from individual psychological aspects
- it is possible to observe objective development of a social system
- it is possible to study the influence of ethical criteria of choice on the social behavior
- it is possible to ask the question concerning what objective aspects of ethics can be studied
- prospectively it is possible to study how different ethical rules evolve and the way they result from the state of the "world" and the way they influence on it.

Such evolutionary approach seems to be the only rational way to evaluate intersubjective rules of ethics.

2. Distributed Artificial Intelligence (DAI)

From philosophical perspective DAI is a cognitive system in which knowledge is possession of a whole system (population of individuals in the case of MAS – Multi Agent Systems). An individual is merely a carrier of such knowledge (or a part of it). Such approach lets us treat the agents in the way similar to biological individuals, which are subject to natural selection. Only two possibilities are open to an agent: it may die or survive in confrontation with its environment, which consists of two components: the "social" (other agents) and the "material" (the world, in which the agent acts). The real influence of selection rules on knowledge is seen only on the level of the population, where strategies are made temporarily stable. On this level advantageous features are strengthened and disadvantageous mutations are eliminated. Such evolutionary game takes place between the whole society (population) and the environment. Thus the way the population acts is more important than the fate of an individual. Intelligence as the adaptive means should reveal this feature. As Weiß [3], one of the prominent persons in DAI research, points out: "intelligence is deeply and inevitably connected with interaction".

The practical realization of computer simulation of MAS looks as follows:

An agent is a part of computer code (program), which is able to act individually within its environment. There is no general definition of an agent but it is assumed that it has to possess learning capability, ability to create an internal model of its environment and the ability to choose the strategy of acting within its environment, based upon a certain (for the time being, a priori given, in the future perhaps self-modifiable) system of goals.

An agent may act socially, creating temporal or stable structures of cooperation with other agents, which may result in specialization in certain actions. This may lead to more complex structures – aggregates of agents – for which the question of self-organization arises. This is not, however, the case of suggested simulation, although the problem of emerging of such structures is interesting in itself.

The agent has two fundamental goals: to survive and to reproduce. The process of agent's reproduction may be enriched by the *crossing-over* mechanism which allows for faster changes and thus increases the population's flexibility.

The agent has to be able to recognize other agents especially those which belong to his own population as potential partners for cooperation.

The selection mechanisms act on individual agents through mechanisms of strengthening and elimination of "unsuccessful" agents. The observations, however, are concentrated on the whole population rather than on individual agents. The intelligence of the system is analyzed in the light of effectiveness of the population in performing the desired actions. This is independent of the goals of individual agents, which are written into agents' "profiles".

One thing concerning the observations is also important. The suggested simulation lets us avoid a certain question about the position of the observer, who in the case of human societies is just a member of the society it examines. In the simulated population the place of the observer is external and, in a way, God-like. He/she is maybe not almighty but able to posses the complete knowledge of the system.

The more detailed discussion of the agent's profile is presented in the parts 4, 5 and 6 of the present paper. Let us only state here that suggested by some authors

(Tuomela [4]) *joint intentions* or any other attempts to implement intentionality or self-consciousness into the agent's profile lie far beyond the scope of our interest. There are two fundamental reasons. Firstly, we are not interested in getting the "human-like" society – with all psychological colorfulness of its participants – but rather the society that acts according to clear rules resulting from what philosophers are able to say about scientific cognition and ethics. Secondly, the evolutionary approach allows for treatment of psychological factors as irrelevant to the knowledge and the rules of acting, which are gained by the whole population.

3. Methodology – why simulate ethical behaviors?

The motivation for simulating ethics in computer environment comes from two reasons. The first lies in computer practice where safety of a computer network depends upon the ability to meet the challenge produced by various negative actions taken by the users (like hackers) and the programs (viruses for instance). The other is purely cognitive. Computer simulations offer a clean (fully controlled) environment, in which it is possible to test certain philosophical concepts in their formal "unpolluted" (for instance by psychological factors) form. Of course one would claim that purely formal, logical analysis should be enough in such cases. This, however, is not true for the concepts based upon evolutionary approach, where purely deterministic predictions are not possible. The world of culture evolution is complex. Studying evolution requires observing multiple generations, which in the case of real (human) society is impossible for purely biological reasons. This, however, does not apply to the societies simulated in computer environment.

Simulation requires simplification. The agents are able to learn and observe their environment and produce the world view. They possess knowledge, or being more precise, their beliefs are identical to the available information about the objects. This knowledge is personal but objective and it is a basis for decision taking. However, until now, the agents used to possess only intellectual capability (intellectual profile) – the ability to rationally analyze the situation. We suggest implementing two additional "profiles" – social and ethical – the former being responsible for the observations of the social environment, the latter corresponding to the rules governing evaluation of what is good or bad. Let us, notice that the core evaluation of being good or bad refers to the observable internal states of an "agent". Those states determine the agent's fitness to the environment thus deciding on the agent's chance to survive.

4. Ethically-social approach to behavior estimation of an autonomic agent on the ground of M-agent architecture

The proposed way to develop intelligent machines by creating "social" machines is the closest to reality if one assumes that intelligence resulted from the evolution and not by miraculous act of creation. The concept is based on the socio-biological theory that primate intelligence first evolved because of the need to deal with social interactions. The model is based on M-agent architecture introduced in [5], [6].

The main idea of the ethically-social estimation of an autonomic agent's behavior is to provide the agent with two additional profiles: the ethical and the social. The ethical profile would cover operations that aim at:

construction of „notion" about oneself on basis of one's ethical rules,

fulfillment of one's ethical rules,

updating and learning the ethical rules on the basis of the environment observation, particularly of those environment elements which are agents.

The group of goals, decisions, operations and knowledge relating to social live mechanisms is named the social profile. This will perform the operations that aim at:

fulfillment of living together rules in society,

observation of environment, especially agents,

„punishing" the agents, that do not fulfill rules of living together.

The motto, which inspired formulating of such a solution is taken from [7]: "There is no society without morality". Any human being taking decision about his/her action is also influenced by the ethical rules he/she knows. The effects of this action are, however, observed through other individuals in society, who suitably, according to the social norms known to them, react on such behavior. The human being obviously watches those reactions and, if the reactions indicate his „bad" behavior, he changes his own ethical norms. Society has an indirect impact on the ethical system of such individual.

The social and ethical profiles are inseparable and mutually supplementary. Social profiles of agents from the environment are an indirect influence on the ethical profile of the agent concerned, as it was shown in the last paragraph. Only thanks to these two profiles it is available to keep social and moral order in open systems, and thus to create mechanisms similar in safety to the ones well functioning in known societies.

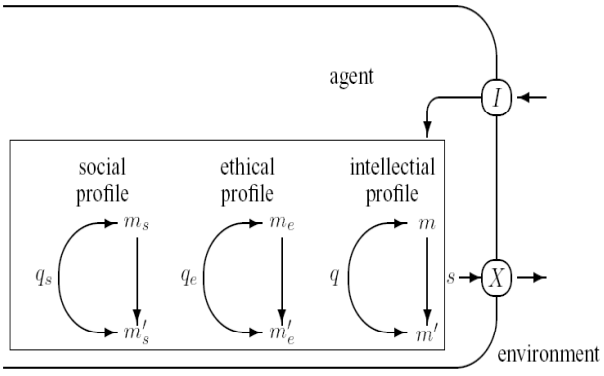


Figure 1. Concept of social, ethical and intellectual profiles in M-agent architecture.

5. Social profile

The social profile forms the class of the agent's activity, which as its goal has the observation of other agents in the society and of other elements of the environment. Those observations should be done in order to distinguish individuals, which do not fulfill social and ethical rules and whose behavior is unfavorable, incorrect or bad according to the observer. The „bad" individuals should be ade-

quately treated (e.g. convicted, avoided) which should be performed by the social profile.

6. Ethical profile

The ethical profile is the class of agent activity, which as its goal has the fulfillment of one's ethical rules and updating or learning the ethical rules on the basis of the environment observation, particularly of those environment elements which are agents.

The operator of agent's ethical learning enables to change the ethical norms in order to be „better" treated in the society.

7. What can be simulated?

Let us first sketch the general scheme of relations between the agent's profiles [see Figure 2]

Here:

The input is firstly dispersed into ethical, rational (intellectual) and social streams.

Ethical, rational (intellectual) and social profiles could be interrelated.

The output is a decision based upon a function of ethical, rational and social worldviews.

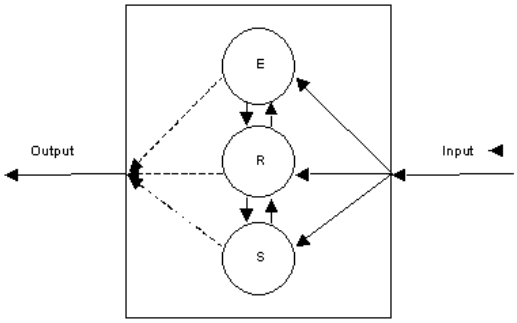
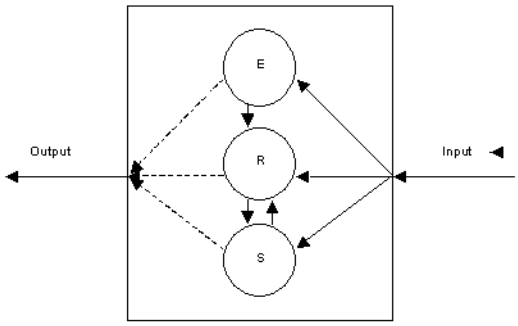


Figure 2. General scheme of relations between agent's profiles

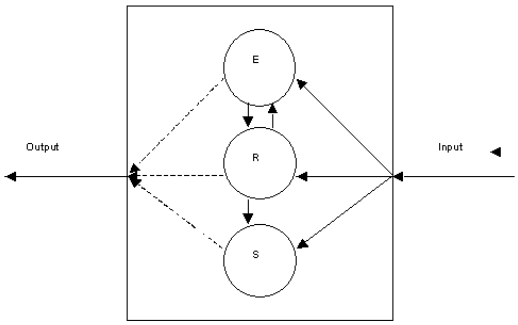
The above scheme must be filled with certain realistic assumptions on various ethical models present in cultural tradition:

7.1 Ethics of moral law



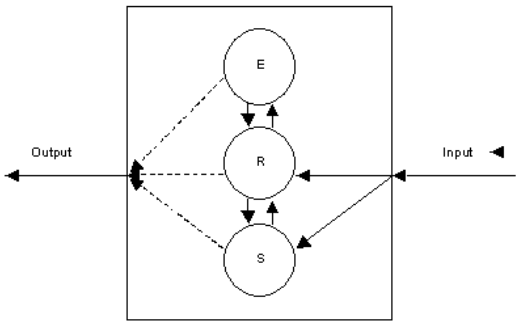
Ethics of moral law is externally given and implemented in the E profile by the programmer and unmodifiable during the evolution. This would correspond to the ethics of natural law. The feedbacks go only from the ethical profile. Let us notice that it allows for simulating the behavior according to various sets of natural laws posing the programmer in God's place.

7.2 Ethics of prize and punishment



Feedbacks here go bilaterally. They let the agent to modify its ethical worldview according to benefits achieved. Possibility of interaction between the social and ethical profiles is noteworthy here. This is, however restricted to the individual fitness of an agent.

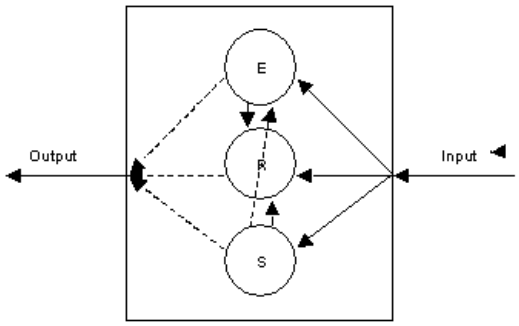
7.3 "Cynics" ethics



The ethical profile is subdued to other profiles, which is realized by implementing the null starting profile. Ethical

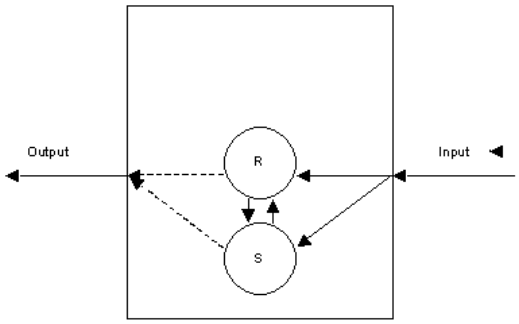
rules are constructed upon the changeable information coming from social and rational profile

7.4 Moral duty ethics



The feedback from the social profile dominates over the one from the ethical profile. The social rules form the shape of the ethical picture of the world, which is independent from rational evaluation

7.5 Null ethics



No ethical profile is implemented or allowed in future.

It is possible to perform simulation of a single model or combine different models in a single simulation, which allows to compare the influence of several or even all profiles on fitness, which gives a competitive situation.

8. What can we learn from such simulations

It is obvious that one should not expect learning anything about actually existing societies[8]. The purity and controllable characteristics of the simulation allows, however, investigating:

the impact of ethics on fitness of social groups,

the differences between the fitness of different models of ethics

the time dependent development of ethics in its social aspect

Concluding one can have the first “empirical” means to study social states in their theoretically (philosophically) constructed forms.

References

Moore G. E. *Principia Ethica* Prometheus, 1988.

Płazowski J. W., Suwara M. *Measuring the Uncertain...",* Proceedings of 25th Wittgenstein Symposium, Kirchberg 2002.

Weiß G. *Ecai-96 workshop on learning in distributed artificial intelligence*. Call for Papers, 1996.

Tuomela, R., 1996, ‘Philosophy and Distributed Artificial Intelligence: The Case of Joint Intention and Joint Action’, in Jennings, N. and O'Hare, G. (eds.), *Foundations of Distributed Artificial Intelligence*, Wiley, New York, pp. 487-503.

Cetnarowicz K. *M-agent architecture based method of development of multiagent systems. Proc. of the 8th Join EPS-APS International Conference on Physics Computing, ACC Cyfronet,* Kraków 1996.

Cetnarowicz K., Nawarecki E., Żabińska M. *M-agent architecture and its application to the agent oriented technology. Proc. of the DAIMAS'97. International workshop: Distributed Artificial Intelligence and Multi-Agent Systems,* St. Petersburg, Russia 1997.

Singer P. *Przewodnik po etyce*, Wydawnictwo „Książka i Wiedza”, Warszawa 1998, 2002.

Płazowski J.W., Suwara M. *Modelowanie Metafizyki* Prace V Krajowej Konferencji “Modelowanie Cybernetyczne Systemów Biologicznych” Kraków 2000.