Social Learning in Population-based Adaptive Systems

E. Haasdijk, P. Vogt and A.E. Eiben

***Abstract*— The subject of the present investigation is Population-based Adaptive Systems (PAS), as implemented in the NEW TIES platform. In many existing PASs two adaptation mechanisms are combined, (non-Lamarckian) evolution and individual learning, inevitably raising the issue of 'forgetful populations': individually learned knowledge disappears when the individual that learned it dies. We propose social learning by explicit knowledge transfer to overcome this problem. Our mechanism is based on 1) direct communication among agents in the population, 2) messages carrying rules that the sender agent uses in its controller, and 3) the ability of the recipient agent to incorporate foreign rules into its controller. Thus, knowledge can be disseminated and multiplied within the same generation, making the population a knowledge reservoir for individually acquired knowledge. We present an initial assessment of this idea and show that this social mechanism is capable of efficiently distributing knowledge and improving the performance of the population.**

I. Introduction

Social learning entails agents learning new skills as the result of interaction with other agents. In this paper we investigate the effects of social learning on a population of autonomous, (virtually) embodied and situated agents. This work fits in the framework of Population-based Adaptive Systems (PAS) as described in [1]. In the most general case such PASs feature evolutionary, individual, and social learning—threefold adaptation, yielding a so-called AAA-PAS. In the present study we investigate the latter of these mechanisms –social learning– in isolation.

In the PAS we examine, agents decide autonomically on the actions they perform by means of a controller that is inherited (for the initial population: generated) at birth. They implement reinforcement learning for individual learning (IL) as well as Evolutionary Learning (EL). Through EL, only the inherited controller is passed on (i.e. *non-Lamarckian* evolution [2]): agents do not inherit knowledge (modifications to the controller) that their parents may have gained through experience; they can only recombine the controllers that their parents had at birth (with some mutation added). This means that, without some additional method of spreading the knowledge through the population of agents, everything an agent learns through experience (i.e., through IL) will be lost when the agent dies.

This is where social learning (SL) comes into play: with SL in place, anything an agent learns during its lifetime can be taught to other agents, so that this knowledge does not necessarily die with the agent that originally discovered it. With agents exchanging knowledge pieces through SL, the population as a whole effectively becomes a knowledge repository –although not a randomly accessible one for individual agents– for IL-discovered knowledge pieces. Obviously, SL can also speed up the learning process at the population level as found in e.g., [3], [4], [5]

Research question

Obviously, SL can only play this role if it can effectively disseminate individually acquired knowledge pieces. The question, then, that this paper seeks to answer is the following:

> Is SL –specifically as implemented in NEW TIES– an efficient mechanism to spread knowledge pieces through the population, thus creating a knowledge repository for individually acquired knowledge?

In nature, SL can be achieved through a host of mechanisms ranging from imitation to social guidance in individual learning [3]. Here, we consider the case where SL consists of agents actively suggesting behavioural rules (knowledge pieces) for the consideration of other agents in a peer-to-peer fashion. The recipients of these knowledge pieces then choose whether or not to integrate them into their own set of rules. The fact that all agents participate in SL at an equal footing implies an inherent parallelism in the spreading of knowledge pieces: all agents that have acquired a knowledge piece can simultaneously share it with other agents, who can then share it in turn, and so on.

Cultural algorithms employ belief spaces [6], which can be seen as explicit knowledge repositories that the individuals build collectively. In the research presented in this paper, however, knowledge repositories are formed implicitly by the population and any individual agent can use only the part of the repository that it embodies. It has been shown that SL through imitation (sometimes called 'cultural evolution') can be beneficial by decreasing the learning time for individuals, particularly in cases where the required behavioural rules are difficult to acquire [3], [4]. Such implementations of SL focus on a limited number of 'experienced' individuals instructing uninitiated individuals one by one and thus do not exploit the inherently parallel ink-stain effect present in the peer-to-peer knowledge exchange that we envisage. Similarly, in ensembles of learning classifier systems, SL –termed 'rule-sharing'– has proved to boost the learning speed [5] of the ensemble. Comparing such ensembles with a population of interacting, mortal agents is tenuous, however: the constituent parts of the ensembles are not considered separately, only the performance of the ensemble's aggregated behaviour is taken into account.

Communication and Information Sciences, Tilburg University, The Netherlands and Artificial Intelligence, Vrije Universiteit Amsterdam, The Netherlands. e.haasdijk@few.vu.nl, p.a.vogt@uvt.nl, gusz@few.vu.nl

978-1-4244-1823-7/08/$25.00©2008 IEEE

Our investigations are part of the NEW TIES[1] project, which is presented in section II. Section III describes our approach to designing and implementing social learning within the NEW TIES project. Sections IV and V present the experiments and their results, respectively. Section VI concludes the paper.

II. NEW TIES

The NEW TIES project aims to create a simulation platform in which a cultural society develops through evolution, individual learning and social learning of autonomous agents [7] . The artificial, virtually embodied agents that make up this artificial society live in a grid world containing objects such as food sources (plants), tokens, places and building bricks.

In this world, time passes in discrete steps. Every time-step, the agents receive stimuli regarding objects (including agents) that they see or carry, messages from other agents that they hear and their internal state (e.g., their own energy level). The agents process these stimuli to select actions such as move or turn, pick up or put down objects, eat, communicate or interact otherwise with other agents (e.g., mating, or giving or taking objects to/from other agents). To process these inputs and arrive at a decision about which action to take, the agents use their individual controllers.

The project models agents anthropomorphically, thereby imposing strict autonomy, (virtual) embodiment and situatedness. This limits our options when designing agent interactions (e.g., agents cannot communicate unless they are within each other's vicinity), perception (e.g., they cannot see inside each other's heads) and learning mechanisms (e.g., no supervised learning).

Decision Q-Trees

The agents' controllers are implemented as a special kind of decision trees, *decision Q-trees* (DQTs). The 'Q' refers to the fact that they can be adapted through Q-learning [8], the NEW TIES implementation of IL. With crossover and mutation operators inspired by those used in genetic programming [9], these trees can also be adapted through EL when two agents mate to create offspring. The following section describes in detail how DQTs can be adapted using the subject of this paper, SL.

DQTs consist of test, bias and action nodes (Fig. 1; depicted as lozenges, trapezoids and rounded rectangles, respectively). Test nodes ascertain whether the required concept is present in the current stimuli. If the test succeeds, the agent traverses to the next node in the left branch; otherwise it traverses to the next node in the right branch.

To traverse a bias node, the agent probabilistically selects one of multiple branches for further traversal – each of these branches has a bias that determines the likelihood of it being selected. These biases are determined genetically through evolution and onto-genetically through individual and social learning.

[1]New Emerging World models Through Individual, Evolutionary and Social learning. http://www.new-ties.org.

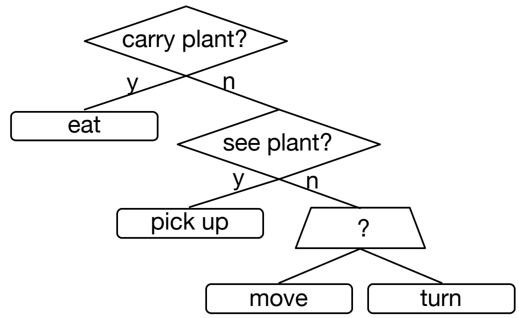

Fig. 1. A simplified example of a decision Q-tree (DQT).

The leaves of the DQT are action nodes that select an action. Action nodes, like bias nodes, are probabilistic: the actual action is stochastically chosen according to a weight distribution over all possible actions. The available actions are simple actions –such as move, turn-left or turn-right–, unary –such as eat(x) or hit(y)–, and binary actions –such as give(a,o)–. The arguments for the higher arity actions are implied by the tests that were traversed to select an action –e.g., testing for visible agents implicitly selects all agents in sight– and can be any object, but if, e.g., an agent attempts to eat a non-food item, this action will fail in the world.

Energy and agent quality

Agents have to husband their energy level: performing the selected action, even if that amounts to inactively surviving a time-step, costs energy. Should an agent run out of energy, it dies. To gain energy, an agent must eat food (plants). Other than that, agents die when they reach a certain maximum age. There is no other selection mechanism: as long as an agent lives, it can act—and therefore, engage in mating or SL. Thus, contrary to typical evolutionary algorithm or evolutionary robotics applications [10], [11], NEW TIES lacks a crisp optimisation criterion as well as a concrete task to be performed optimally. The agents survive whatever the environment throws at them or they do not—that's all there is. This also entails that there is no measure of fitness in this system: the only selection mechanism is –truly Darwinian– the struggle for life in the environment: environmental selection.

To gauge their relative quality, agents can, however, be compared in terms of their perceivable attributes such as age or energy level. Crucially, such comparisons aren't performed by some central selection mechanism –as would be the case in traditional evolutionary algorithms–, but by the individual agents themselves when they *autonomously* decide to mate, engage in SL, or otherwise interact with another agent.

III. SOCIAL LEARNING

We have chosen for a push model, where teachers volunteer knowledge pieces that the students then may accept. The NEW TIES platform does not impose this design choice; it can just as well accommodate a pull model, where agents request knowledge from other agents. A combined model, where agents advertise that they believe that they have useful knowledge to share and other agents can then request that knowledge (similar to the plumage concept in [12]) could be implemented in NEW TIES as well. Similarly, we have implemented a measure of relative quality $R(a,b)$ (described below) that compares agents a and b in terms of energy and age, but we could have chosen, for example, a reputation-based measure as well.

Generally, this section describes the current implementation of SL within NEW TIES—alternative design choices could be made and implemented at every level described here. As mentioned above, however, some options are unfeasible because of anthropomorphical nature of NEW TIES agents. For instance, agents have to be within range ('earshot,' if you will) to be able to communicate and hence engage in SL.

SL is implemented in the following sequence for every agent at every time-step:

1) An agent chooses to initiate sending ('teaching') probabilistically ($p = 0.2$).
2) If it decides to send, the agent describes the trace through its DQT that led to the current action (e.g., "I'm moving because there is no food to pick up").
3) Of all the agents in range, the teacher then selects the one with the lowest energy as the 'student'.
4) When an agent receives a knowledge piece, it stochastically chooses to integrate ($p = 0.2$) or disregard it.
5) When an agent s incorporates a DQT path P it received from an agent t, agent s selects the most similar path P' in its own DQT according to the following criteria:
   a) Percentage of matching tests
   b) The number of tests P but not in P'
   c) The number of tests in P' but not in P

   If the percentage of matching elements in P is 100%, the bias for the action that P results in is multiplied with the relative quality $R(t,s)$ (see below). Otherwise, the agent engages in a kind of dialectics: it inserts a bias node at the first point of divergence between P and P'. The remainder of P' is inserted as one option at that node, a sub-tree corresponding to the non-matching entries in P is inserted as the alternative. The biases for the options are set proportionally to the relative quality $R(t,s)$. Figure 2 illustrates this procedure.

As described above, our SL implementation requires some measure of (relative) quality for agents to be able to assess the merit of received knowledge pieces when incorporating those pieces. To that end, an agent a can determine the relative quality $R(a,b)$ of another agent b from their relative ages A_a and A_b and energy levels E_a and E_b, respectively:

$$R(a,b) = 0.5 \cdot (\frac{A_a}{A_a + A_b} + \frac{E_a}{E_a + E_b})$$

This measure ranges from 0, where agent b devastatingly outperforms agent a to 1, where the converse is true. If the agents have the same energy and are equally old, $R(a,b)$ equals 0.5. Note that this measure does not constitute an optimisation criterion as typically used in evolutionary algorithms: it does –without specifying any goal– allow for the comparison of the success of adaptation of individuals.

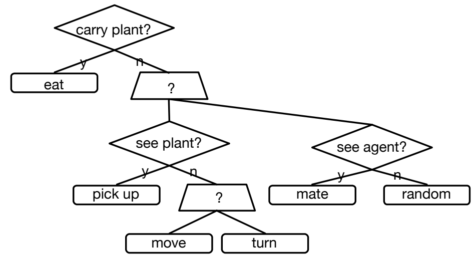

Fig. 2. The result of integrating the path `[not carry plant; see agent]` $\Rightarrow$ `mate` into the DQT from Fig. 1.

SL as an Evolutionary Algorithm

[12] already showed that an agent-based knowledge exchange mechanism similar to SL constitutes an evolutionary algorithm (EA). Moreover, as put in to [10], an EA requires

- Selection as a force to push quality
- Variation operators to create the necessary diversity and thereby create novelty.

This implementation of SL achieves the former of these at various levels. Firstly, ill-adapted individuals tend to die relatively quickly, and hence cannot further distribute their knowledge, while well-adapted individuals tend to survive and have ample opportunities to distribute their knowledge. The second level is that of student selection mentioned above: when an agent has to choose between potential recipients of a knowledge piece, it selects the one with the lowest energy. Finally, the integration mechanism uses the relative quality $R(a,b)$ to set the bias for already known or newly received knowledge.

Variation is provided by the knowledge integration mechanism, which can be seen as a guided adaptation of crossover such as commonly used in genetic programming. Although this suffices, IL and SL dovetail very nicely in this respect (as well as because of the benefit that we expect from SL providing a knowledge repository for IL): IL then plays the part of a mutation-like variation mechanism.

IV. Experimental set-up

As noted above, the system is not meant to set the agents any specific task other than to win the struggle for life. The environment can, of course, be set up to challenge the agents in specific ways. The agents then have to deal with these challenges in order to survive.

In the experiments we describe here, the environment is set up such that agents can only survive if they successfully tackle the well-known poisonous food problem [13], [14],

[15]. The agents find themselves in an environment where there are two types of plants, both of which can be picked up and eaten. One type is nutritious and yields an energy increase, the other type is poisonous and eating them actually drains energy. Agents can choose not to, but they can distinguish between the two types of plant. They do not, however, know a priori that one kind –let alone which kind– is poisonous. Because agents *must* eat to replenish their energy level as mentioned above, they have to learn to disregard poisonous food if they are to survive.

To measure the efficacy of SL as a mechanism for the proliferation of knowledge pieces through a population, we ran a series of experiments where the population consists of two kinds of agents: knowers and students. The knowers have pre-built controllers that allow them to tackle the poisonous food problem. The students have a partially randomly constructed controller—they know how to pick and eat plants (regardless of their being poisonous or not) and how to mate, but the rest of their DQTs is constructed randomly. A varying proportion of the agents with pre-built controllers can send, but not receive SL messages ('teachers'), while students both send and receive SL messages. The remaining knowers do not engage in SL in any way; they are only there to ensure that the environment contains the same amount of agents eating away at the wholesome plants across the experiments, so that the results are comparable.

Another difference between students and knowers is that the former can mate to produce offspring where the latter cannot. Note, that this does not –in these particular experiments– constitute evolution: there is no variation operator because it does not entail recombination, but cloning of either parent. Therefore, there is no evolutionary learning at play to disturb our measurements. Neither kind of agent can perform IL in these experiments.

This set-up serves as an idealised exemplar of a population where some agents –represented by the teachers– have discovered, through individual learning or otherwise, a particularly useful bit of knowledge: to eat only wholesome plants. Note, that these teachers play quite a different role from the 'experienced individuals' employed by [3], [4]: from the students' point of view, they are no different from any other agent they encounter. We ran the experiment with varying numbers of teachers to compare the rate at with which the population of students learns to differentiate between nutritious and poisonous food.

In our experiments, the agents can move in a 200×200 grid. There are initially 250 students and 100 knowers, of which 0, 1, 5 or 50 individuals are teachers. Agents can live well beyond the length of the experiments, so agents can only die of lack of energy. Each experiment was repeated 20 times. Poisonous plants drain 1.5 times the energy that wholesome plants yield, the environment is initialised with 16,000 plants of each type. Plants regrow practically immediately (within 2 time-steps), even if they've been picked, similar to food in SugarScape [16]. Thus, there is always food (and poison) available and the ratio poisonous/wholesome plants more or less remains at the initial value of 0.5.

Measurements

As a behavioural measure, specifically for the poisonous food environment we use a function based on the ratio of the different types of food the students eat:

$$g(t) = \frac{\sum_{t-1}^{t} eat_h}{\sum_{t-1}^{t} eat_p + \sum_{t-1}^{t} eat_h}$$

Where $\sum_{t-1}^{t} eat_h$ and $\sum_{t-1}^{t} eat_p$ are the number of wholesome and poisonous plants eaten by the population between $t-1$ and t.

We also employed a structural measure that actually detects the presence of the required knowledge. There are, of course, many different strategies that allow the agents to eat only wholesome plants—e.g., "only pick up wholesome plants and eat anything you carry", or "drop any poisonous plant and eat anything you still carry". In these experiments, however, we know exactly which knowledge piece we expect to find because it is the relevant trace through the handcrafted knowers' DQT: it's `[carry wholesome plant]` $\Rightarrow$ `eat`. This allows us to identify, during a run, those students that have incorporated this rule by inspecting their DQTs. Thus, we can measure the incidence among the students of the appropriate knowledge piece.

Note, that the measurements we present here were taken only over the population of students.

V. Results

Figure 3 shows the development over time of $g(t)$ –averaged over 20 runs– for the students with 0, 1, 5 and 50 teachers. For reasons of legibility we omitted error bars; the 4 curves do differ significantly, although the standard deviation for 0 and 1 teacher is large, due to the fact that in many of these simulations, the students didn't eat at all.

As can be seen, $g(t)$ remains level just above 0.5 for 0 teachers –there is no learning at all– the slight improvement over fully random behaviour is due to environmental selection: agents that eat too much poisonous food simply die at a faster than agents that do not or less so, leaving a slightly better set of surviving agents. In the case with a single teacher, the performance of the students increases significantly: even from so small a seed, a knowledge repository can grow. For 5 and 50 teachers, the population behaviour improves rapidly until $g(t)$ reaches a plateau between 0.8 and 0.9—there is no significant difference between these experiments after that point. This seems to imply that in both cases the population of students becomes saturated –at least at a behavioural level–with the appropriate knowledge piece.

Figure 4 shows a series of maps of the world displaying the incidence of the required knowledge piece (`[carry wholesome plant]` $\Rightarrow$ `eat`) geographically. The three sequences of maps show the spread of knowledge over time for typical runs with 1, 5 and 50 teachers respectively. Students that contain the required knowledge show white,

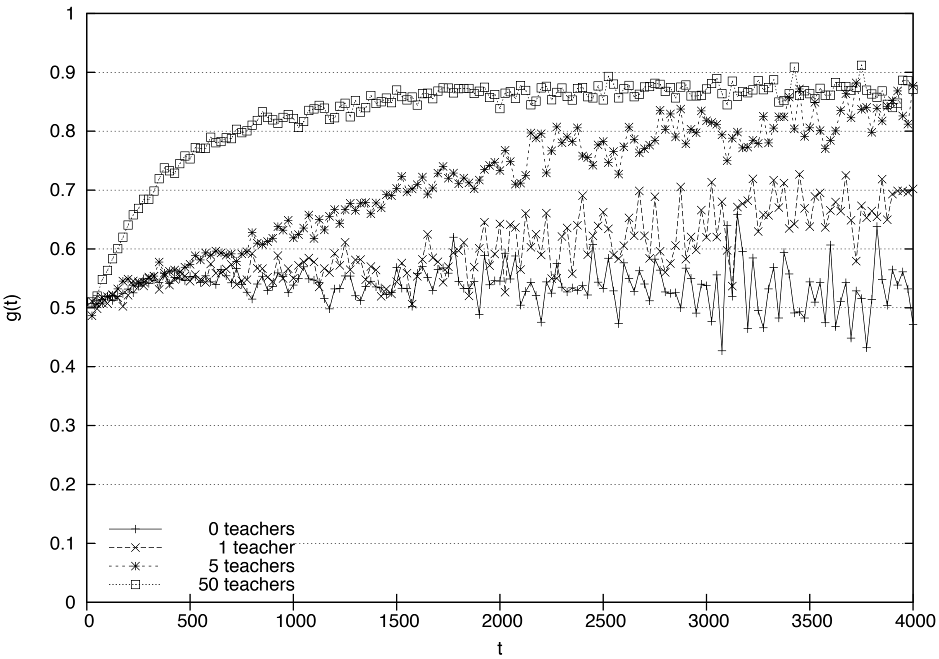

Fig. 3. Development over time of $g(t)$ for the student population for different numbers of teachers.

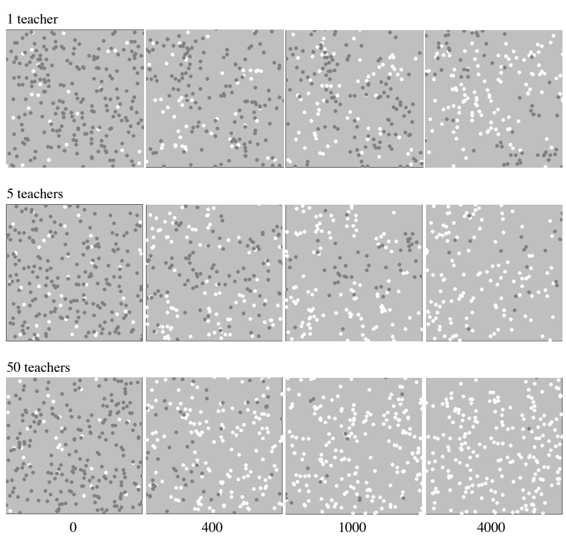

Fig. 4. Spread of knowledge pieces over the students for typical runs with 1, 5 and 50 teachers at timesteps 0, 400, 1000 and 4000.

those that don't show dark grey. Teachers and knowers are not shown. Note the logarithmic time-scale.

Again, it is plain that, even with a single teacher to initiate dissemination, the decisive knowledge is spread through a significant part of the population—the population as a whole stores the knowledge effectively and robustly. As could be expected, the knowledge becomes even more widespread for the experiments with 5 and 50 teachers.

While we have seen the behaviour for the student population reach similar levels for the experiments with 5 and 50 teachers, this is not the case for the incidence of the expected knowledge piece. With 50 teachers, practically all students have obtained this knowledge piece after 4000 time-steps, but in the 5 teachers case, a portion of the students remains unaware of this information at that time. Similarly, there is no significant difference between $g(t)$ at time-step 1000 and at time-step 4000 for the 50 teachers experiments, but there is a marked difference in incidence of the required knowledge piece. From this we can conclude that, after a certain level of prevalence has been achieved, further proliferation of the knowledge piece has has no perceivable effect on population behaviour in terms of $g(t)$.

Figure 5 shows how the percentage of students that have learned the requisite knowledge develops over time with 1, 5 and 50 teachers, respectively, averaged over 20 runs. Because

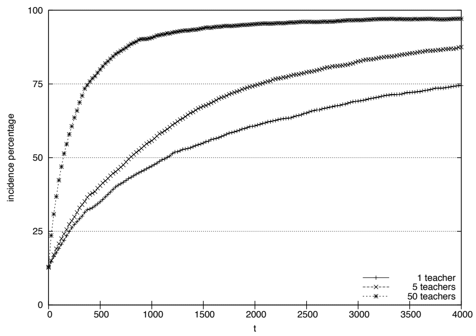

Fig. 5. Development over time of the percentage of students with the crucial knowledge piece.

the students spread the knowledge they receive, incidence grows almost exponentially as can be seen from the graph.

Note, that at time-step 0, a portion of the population does contain the knowledge as part of the randomly initialised tree while $g(t)$ for the runs without any teachers doesn't increase over time. This can be explained by the context in which the knowledge piece may be present (i.e., as sub-clause in a more complex, possibly nonsensical rule) and by the fact that the action node's weights (as described in section II) aren't sufficiently biased towards actually selecting the `eat` action.

VI. Conclusions and future work

In this paper we presented SL as implemented in NEW TIES. We investigated whether SL can provide a successful mechanism to spread knowledge pieces over a population, a minimum requirement to enable the population to create a knowledge repository for otherwise volatile individually acquired knowledge.

To this end, we ran experiments with an environment that poses a straightforward challenge –the poisonous food problem– to the agents. We devised an idealised situation where teachers were introduced to play the part of experienced individuals who have acquired beneficial knowledge. The teachers were introduced into an environment of students. We measured how successfully the students tackle the challenge as well as how the crucial knowledge piece spread through the population.

The results clearly show that SL is an efficient mechanism for the dissemination of knowledge pieces through a population of agents. Even from a single agent, the knowledge can spread over the majority of the population like an ink-stain on tissue paper. Within the framework of AAA-PAS in general and the implementation in NEW TIES in particular, this means that SL is capable of allowing the population to form a knowledge repository for individually acquired knowledge so that such knowledge doesn't necessarily expire with the agent that discovered it.

After considering SL by itself, it stands to reason to investigate its impact when combined with the other two As in AAA-PAS: EL and IL. We dwelt on the synergy between SL and IL on a number of occasions above; it will therefore come as no surprise that we intend to investigate this particular combination of learning methods especially. In parallel, we are investigating the impact of using agent-acquired language as the medium to exchange knowledge in SL in stead of the explicit meaning transfer we used here [17].

Acknowledgements

The NEW TIES project, of which this work is part, is supported by an European Commission FET grant under contract FP6-502386. Information provided in this manuscript is entirely the authors' responsibility; it does not reflect the views of the Commission or even necessarily that of other NEW TIES members. We thank all project members for their invaluable contributions and suggestions. Also, we would like to thank Zoltán Papp, Martijn Schut and Selmar Smit for their contributions.

References

[1] A. E. Eiben, A. R. Griffioen, and E. Haasdijk, "Population-based adaptive systems: concepts, issues, and the platform new ties," in *Proceedings of ECCS 2007*, 2007.

[2] J. B. Lamarck, *Philosophie zoologique, ou Exposition des considérations relatives à l'histoire naturelle des animaux*. H.R. Engelmann, 1809.

[3] A. Acerbi and S. Nolfi, "Social learning and cultural evolution in embodied and situated agents," in *Proceedings of the First IEEE Symposium on Artificial Life*. Piscataway, NJ: IEEE Press, 2007, to appear.

[4] D. Denaro and D. Parisi, "Cultural evolution in a population of neural networks," in *Proceedings of the 8th Italian Workshop on Neural Nets*. London: Springer-Verlag, 1996, pp. 100–111.

[5] L. Bull, M. Studley, A. Bagnall, and I. Whitley, "Learning Classifier System Ensembles With Rule-Sharing," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 4, pp. 496–502, August 2007.

[6] R. G. Reynolds, "Cultural algorithms: Theory and applications," in *New Ideas in Optimization*, D. Corne, M. Dorigo, and F. Glover, Eds. London: McGraw-Hill, 1999, pp. 367–377.

[7] N. Gilbert, M. den Besten, A. Bontovics, B. Craenen, F. Divina, A. E. Eiben, A. R. Griffioen, G. Hévézi, A. Lörincz, B. Paechter, S. Schuster, M. Schut, C. Tzolov, P. Vogt, and L. Yang, "Emerging artificial societies through learning," *Journal of Artificial Societies and Social Simulation*, vol. 9(2), 2006.

[8] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press, 1998.

[9] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: The MIT Press, 1992.

[10] A. E. Eiben and J. Smith, *Introduction to Evolutionary Computing*. London: "Springer-Verlag", 2003.

[11] S. Nolfi and F. Floreano, *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. Cambridge, MA. / London: The MIT Press, 2000.

[12] R. Smith, C. Bonacina, P. Kearney, and W. Merlat, "Embodiment of Evolutionary Computation in General Agents," *Evolutionary Computation*, vol. 8, no. 4, pp. 475–493, 2000.

[13] A. Cangelosi and D. Parisi, "The emergence of "language" in an evolving population of neural networks," *Connection Science*, vol. 10, pp. 83–93, 1998.

[14] S. Nolfi and D. Parisi, "Learning to adapt to changing environments in evolving neural networks," Institute of Psychology, National Research Council, Rome, Italy, Tech. Rep. 95-15, 1995. [Online]. Available: citeseer.ist.psu.edu/nolfi96learning.html

[15] P. M. Todd and G. F. Miller, "Exploring adaptive agency ii: simulating the evolution of associative learning," in *Proceedings of the first international conference on simulation of adaptive behavior on From animals to animats*. Cambridge, MA: The MIT Press, 1990, pp. 306–315.

[16] J. M. Epstein and R. Axtell, *Growing artificial societies: social science from the bottom up*. Cambridge, MA.: MIT Press, 1996.

[17] P. Vogt and E. Haasdijk, "Social learning of skills and language," in *Proceedings of Int. Workshop on Social learning in embodied agents*, A. Acerbi, D. Marocco, and P. Vogt, Eds., 2007.