Selection of Behavior in Social Situations

Application to the Development of Coordinated Movements

Samuel Delepoulle[1 2 *], Philippe Preux[2], and Jean-Claude Darcheville[1]

[1] Unité de Recherche sur l'Évolution des Comportements et des Apprentissages (URECA), UPRES-EA 1059, Université de Lille 3, B.P. 149, 59653 Villeneuve d'Ascq Cedex, France, lastname@univ-lille3.fr

[2] Laboratoire d'Informatique du Littoral (LIL), Université du Littoral Côte d'Opale, UPRES, B.P. 719, 62228 Calais Cedex, France, lastname@lil.univ-littoral.fr

Abstract The law of effect is a very simple law which relates the probability of emission of a behavior by a living being to the consequences of the emission of this behavior by this living being in the past. As such, this law models very basic learning. This law can be considered as an experimental fact as far as it has been observed for a whole range of living beings including human beings. In this paper, we first show that this general law can be the result of a selection process such as natural selection. Then, we show that the implementation of this law can lead to the design of adaptive systems which can mimic very closely the way a new-born develops coordinated movements. To sum-up, we show that the ability to learn such coordinated movements and exhibit adaptive behaviors can result from a multi-stage process of selection.

1 Introduction

The selection of behaviors by their consequences is of paramount importance in the study of behavior of living beings among behavior psychologists, and ethologists. Initially introduced by Thorndike as the "law of effect" [14,15], this principle has been put to the test, verified, and reported in an innumerable amount of situations and publications. The law of effect holds for all living beings, and it has been investigated for creatures ranging from flies to human beings. The law of effect solely deals with learning during lifetime: hence, it deals with the evolution of behavior of a living being (that is, the evolution of its behavioral repertoire) during its lifetime. We want it to be clear that the law of effect has nothing to do with natural selection proposed by Darwin and successors. Natural selection acts along generations on populations of living beings; selection of behaviors by their consequences and its model, the law of effect, acts along lifetime among populations of behaviors. Given a living being, natural selection has produced innate abilities, while selection of behaviors by

* during this work, Samuel Delepoulle acknowledges the support from "Conseil Régional Nord-Pas de Calais, France", under contract n 97 53 0283

their consequences produces its acquired, or learnt, abilities. In the following, we consider that natural selection is well-known by the reader but we will describe selection of behaviors by their consequences in more details.

The law of effect merely states that when a living being emits a behavior that brings it back favorable consequences, the probability that the living being emits this behavior again in the "same" situation increases. This process leads to a selection of behaviors during the lifetime of an animal. It may seem strange, and even unbearable, to state that human behaviors (at least, some of them) are driven by such a simple law. However, two things should be made clear: it is not claimed that all animal behaviors follow this law; it is also clear that so-called "cognitive" activities are taking place in human beings, as well as in most animals. Anyway, some people argue that cognitive activities may also be explained using the law of effect, but this old and important debate is clearly out of the scope of this paper. The fact is that the evolution of complex behaviors of living beings (including human beings) have successfully been explained using the law of effect [5]. Indeed, the interaction of a set of agents which behavior is driven by the law of effect can show complex patterns of activity, and a continuously adaptive behavior. In this regard, adaptation is considered as learning [12]. A nice thing is that the law of effect is a selectionist law, just like natural selection [11]. However, natural selection works along generations on populations of individuals, while the law of effect works on populations of behaviors during lifetime. Clearly, the law of effect may also have indirect consequences on future generations according to the Baldwin effect. Different authors have studied the influence of learning on evolution (see e.g. [1,6,8]). The interaction between the learning process and the evolution process being important, we come back to this issue in Sec. 4. It is also worth noting that the implementation of the law of effect in computer programs is not obvious although it is simple to express verbally. If that was the case, the problem of creating truly adaptive artefacts would probably be solved to a greater extent.

In this contribution, we aim to set a bridge between different results we have recently obtained. This work is the result of a collaboration between psychologists specialized in the experimental analysis of behavior and computer scientists. The line of thought will give the organization of the sequel of this contribution. First, we show that the law of effect may be acquired by way of genetic selection. Second, we show that agents which behaviors are selected according to the law of effect perform well in a social situation. Then, grounded on a parallel between the law of effect and reinforcement learning, we show that it is possible to model in a very realistic way the development of the reaching arm movement in human beings using reinforcement algorithms. While these results are interesting to scientists studying life, they are also interesting for computer scientists to build new adaptive systems that are deeply inspired by available knowledge on the behavior of living beings, the most adaptive systems we are aware of so far.

2 The selection of the law of effect

In this section, we present briefly the result of computer simulations showing that the law of effect can have been selected among generations by natural selection. As far as we are aware of, this important result has never been demonstrated so clearly up to now. This section summarizes chapter 12 of [3].

In this simulation, a population of agents evolves by way of a genetic algorithm. Each agent interacts with its environment via a set of N input sensors, and N behavior units. The input sensors are sampled at each time step. The emission of a behavior is performed by the behavior units ; at each time step, at most one behavior unit can be active, emitting the corresponding behavior. The activity of an agent is coordinated by a neural network. This neural network is made of C layers of N neurons. A neuron of a layer is never connected to a neuron of the same layer. A neuron can be connected to any neuron of the 2 surrounding layers. Each input sensor is connected to all N neurons of the first layer. Each behavior unit is connected to one neuron of the last layer and feeds these N neurons back. The response of each neuron is characterized by 6 real numbers and a boolean value which indicates whether the neuron is active or not. Due to the inter-connection topology, each neuron is also characterized by $2 \times N$ weights. Each weight itself is characterized by a quadruple. Without going into all details which would require much more space, the parameters of a weight describe how the weight evolves along time and activations. Finally, the whole network is characterized by two numbers A_c and A_p. The activation of the neurons of an agent is made at random: A_c indicates the number of neurons that are activated at each time step, while A_p specifies the number of weight updating steps (or, learning step) that are performed by the network. At each time step, a neuron is drawn at random. So, in general, there are less than A_c different neurons that are activated at each time step, while some of the neurons are activated more than once.

An initial population of 10 agents is formed at random, that is that the characteristics of each neuron are drawn at random. Then, the agents are evaluated and follow the usual genetic algorithm loop of selection, duplication, recombination, and mutation. Recombination is one-point crossover which cuts only between two neurons (instead of between any two bits). Five different mutations are used, each with its own probability: mutation of a weight which means changing the value of a weight of $\pm$ at most 10 %; mutation of a neuron which means resetting all the parameters of a neuron at random; mutation of expression which means toggling the activation bit; mutation of A_c or A_p.

The task that has to be performed by the agents is a discrimination task: two stimuli S1 and S2 are presented alternately; in presence of a certain stimulus, the agent must emit behavior B1; in presence of the other stimulus, the agent must emit behavior B2. Each stimulus is input on a given neuron of the input layer. If the agent behavior is correct with regards to the presented stimulus, then it is reinforced: the evaluation of the agent is equal to the number of reinforcers received over 10 sessions, also called its "score". In each session, a certain association has to be learnt: S1-B1, and S2-B2, or S1-B2, and S2-B1. Then, the next session, the other association may become the correct one. 50 % of the sessions reinforce the S1-B1/S2-B2 associations, while the 50 other % of sessions reinforce the other association. No extra-stimulus indicates which association is reinforced in a given session. Furthermore, it should be mentioned that sessions are not known by the agents; for agents, there is no difference except the fact that they do not receive a reward for the same behavior. Each session is made of 1000 presentations of stimulus. This way of changing the reinforced association selects adaptive agents, not only agents that can learn a given association.

Though this description of agents may seem complicated, it is rather simple and natural and seems to be rather minimal with regards to the task we wish to accomplish.

Simulations have been performed. After 200 generations, approximately 90 % of the agents are able to accomplish the discrimination task.

So, this means that a rather simple neuron architecture can be selected along generations to realize a discrimination task. Thus, discrimination abilities in living beings might be the product of evolution. Being able to discriminate, that means that these agents follow the law of effect: their behaviors are selected by their consequences. So, this means that agents whose behavior are selected by their consequences can be selected by selection along generations, that is, by natural selection.

3 The law of effect selects social behaviors

In this section, we show that agents which behaviors are selected by the law of effect outperform by far other agents in a social task, the minimal social situation [10]. This situation is well-known in the social psychology literature. This section summarizes [4].

We do not describe the minimal social situation itself but how we have modeled it to simulate it: this will provide the reader enough information to follow the line of thought. Interested readers are kindly asked to refer to [10] for a thorough description of this situation. We have simulated the realization of this experience by the agents described in the previous section. In this case, the task to be accomplished is as follows: in 50 % of the sessions, if the behavior of agent A is B1, then the agent B receives +1; if A emits B2, then B receives -1; if B emits B1, then A receives +1; if B emits B2, then A receives -1. In the other 50 % of sessions, the consequences of each agent behavior are exchanged: B1 pays -1 to the other agent while B2 pays +1 to the other agent. So, one's behavior only affect his party's reward. The best strategy is for both to emit the behavior that provide +1 to its party. This behavior is a cooperation.

Two sets of simulations have been performed. The only difference between these two sets is the way how the initial population is formed. In the first set S_r, the initial population is formed at random; in the second set S_s, the initial population is formed with agents that have been able to select their behaviors according to their consequences at least some times to times: they are not very good at this task, but they are able to achieve it with statistical significance.

The performance of an agent in the minimal social situation is measured by the number of $+1$ it receives. Figure 1 plots the evolution along the simulation of these two populations. There is a very clear advantage for agents of S_s which obtain good performances after 50 generations, and already much better performances than S_r after only 20 generations.

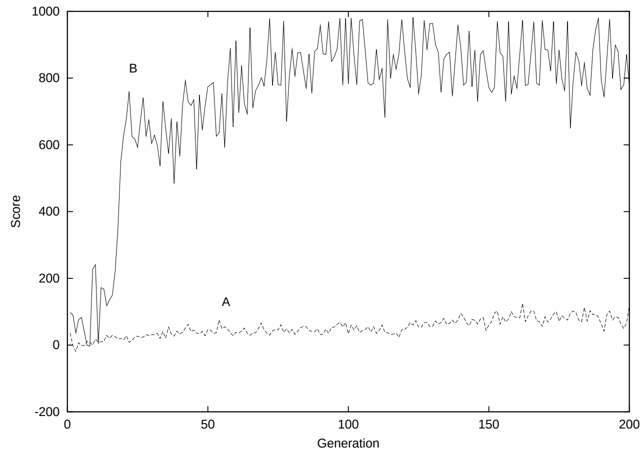


Figure 1. Evolution of the performance of agents that succeed in the minimal social situation. The curve (A) corresponds to the case where the initial population is formed with agents selects their behaviors according to their consequences that at least some times to times. The curve (B) corresponds to the case where the initial population is formed at random. The difference in performance of the two populations is striking. The best possible performance would be rated 1000.

Thus, in this section, we have shown that agents which select their behaviors according to their consequences can show a striking advantage in a social situation which selects behaviors that lead to cooperation. We emphasize the similarities between the dynamics observed here with the population S_s and with human subjects. So, in this situation, the law of effect selects cooperative behaviors.

4 Social behaviors select the dynamics of the arm reaching movement

Up to now, we have shown that the law of evolution of the behavioral repertoire of a living being during its lifetime can be the product of natural selection. So, natural selection (NS) has induced the behavior selection mechanism (BS), that is, the law of effect which yields the selection of behaviors by their consequences along lifetime; this induction is denoted by thick arrows in figure 2. It is known that behavior during lifetime can modify the genetic material of species by the

Baldwin effect [1]; this feedback is denoted by thin arrows in figure 2. At this stage, it is worth noting that due to the law of effect, the behavior of a living being is modified by its environment, among which, the living beings with which it interacts during its lifetime; this interaction is shown with zigzagged arrows between two NS/BS feedback loops, that is, between two living beings in figure 2. Then, we have a sketch of the way different evolution processes interact to yield living being behaviors. It should be clear that these processes happen at different level (genetic/behavior) and different time scales, as well as the interactions between these processes (see figure 2). NS deals with the adaptation of genomes along generations, while BS deals with the adaptation of behavior during lifetime.

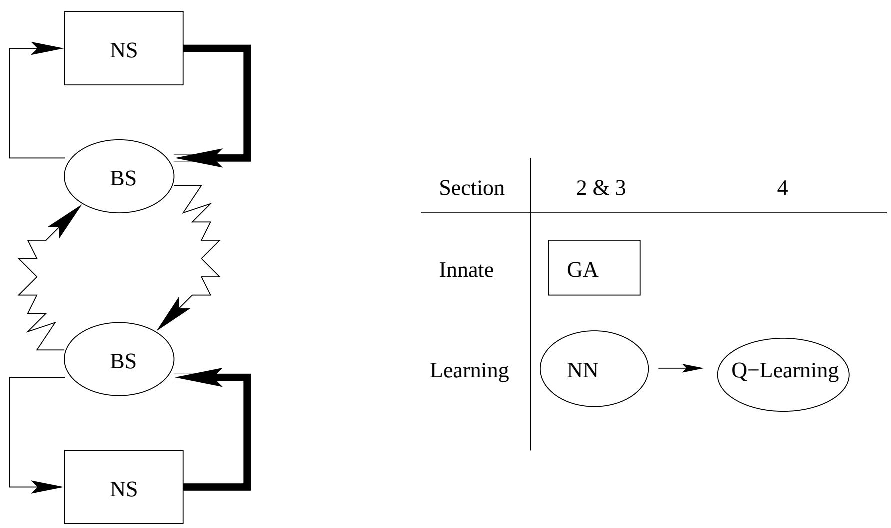


Figure 2. On the left part, we sketch how the natural selection process (NS) interacts with the behavior selection process (BS), as well as, how the behavior of one living being interacts with that of other living beings (zigzagged arrows): the upper part and lower part NS/BS loops each represents one living being, while the zigzagged arrows represent the interaction between living beings. Clearly, all the environment of a given living being may alter its behavior, but we have only represented the interaction between two beings for the sake of clarity. This sketch also shows the interaction between innate and acquired abilities. It is note worthy that the processes happen on different time scales: NS acts along generations while BS acts along lifetime; likewise, the time it takes for one process to alter an other one is variable, ranging from fractions of a second to generations. The right part indicates how the various processes have been implemented in our work. The shape of the boxes in the left part (a rectangle for NS, an ellipse for BS) refers to the table in the right part: NS is implemented with a genetic algorithm (GA) while BS is implemented by way of neural networks (NN) in Sec. 2 and 3, with Q-learning in Sec. 4.

The two processes are implemented by way of two different algorithms: a genetic algorithm implements natural selection while neural networks and Q-learning implement the behavior selection. Natural selection yields the ability to learn to living beings, while behavior selection lets them acquire, and develop their behavioral repertoire, and learn new abilities during its lifetime.

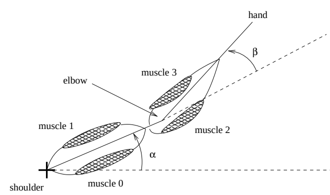
We have used two algorithms (NN, and Q-learning) to implement the behavior selection process. However, it should be stressed that this is a purely technical point: it is only a matter of efficiency in computations and simplicity in software development: we could have used the neural networks of Sec. 3 in place of Q-learning in Sec. 4. The fact is that we are currently updating the simulator so that this is done that way, using NN instead of Q-Learning.

This having been said, we now describe the interaction (during their lifetime) between two agents which behaviors are selected by their consequences.

Lots of organs and members of living beings are made of cooperating elements: muscles, tendons, ligaments, bones, ... not mentioning neurons which play an important role in the overall organization of movements and behaviors in general. In this section, we show that a combination of cooperative agents whose behaviors are selected by their consequences may simulate the development of the arm reaching movement showing remarkable similarities with those of a human baby facing the "same" situation. Instead of using the agents that have been selected in the simulations reported in the previous sections, we use the Q-learning algorithm to model the law of effect. Indeed, Q-learning models rather well the law of effect. This point has been argued at length in [2] and [13]. This section summarizes [9].

Our goal is to model the reaching movement of an arm. We describe here the 2 dimensional case. An arm is modeled as being composed of two segments joint by an articulation (see fig. 3). One extremity of the arm is fixed (the "shoulder") while the other one can move (the "hand"). Each segment is controlled by a set of two muscles: an agonist muscle and an antagonist muscle. So, there are four muscles. Each muscle can be in either 1 of 50 states which indicate its tension. At each time step, this tension can increase or decrease of 1 unit. Increasing or decreasing its tension are the two possible behaviors of each muscle-agent. Each agent receives three inputs: its current state, and two stimuli that provide very poor visual information. Actually, each agent models a muscle and the motor-neuron associated to it. This gives relevancy to the fact that the agent receives visual information. Finally, the behavior of the agent has a cost in energy. The model, though simple, is rather realistic.

Like the human babies with whom the experiment has been done, the arm receives a reward only when it puts its hand into a certain region of the space. Otherwise, it does not get any reward. So, the system of agents has to reach a balance between the cost of emitted behaviors and the need to put its hand in the reinforcement zone to get its reward. The agent is not initially aware of the fact that it can receive any reward. At the beginning of the simulation and as long as its hand does not come into the reinforcement zone, the agent receives no consequence for its behaviors, except their energetic cost. The visual system indicates whether the hand is rather close of far from this region with an integer 0, 1, 2, or 3 and the relative position of the hand with regards to the region: north-west, north-east, south-west, south-east. A very large majority of points of the space are merely perceived as far, in a certain quadrant (see fig. 3). This visual system has been designed to be rather crude. There are two reasons for that: first, in the new-born, the visual system is also very crude. Second, from a computer scientist point of view, if the visual system was very acute, then, it would be very simple to solve the problem we have assigned the arm by simply moving the hand towards the target using a mere gradient algorithm. As far as the vision does not provide enough information to guide the hand towards the target, such a trick is not possible. This is precisely what we wanted to avoid: putting the solution of the problem into the agent.



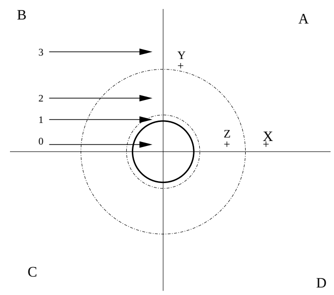


Figure 3. The leftmost part sketches the architecture of the arm (see text for description). The rightmost part indicates how the arm perceives its environment. The visual field is split into 4 quadrants (A, B, C, and D) and 4 circular zones (0, 1, 2, and 3). Positions X and Z are perceived differently; positions X and Y are perceived as identical though more remote than X to Z. So, this visual system is rather crude and does not help very much the arm to reach the reinforcement zone.

The simulation of this arm shows different stages of development of the movement (see fig. 4) : initially, the hand wanders erratically; after some times, it passes into the reinforcement zone, by chance, but it is unable to remain in it; after the position of the arm has been reset to its initial position, and after many attempts, the arm develops a smoother and smoother direct movement to reach the zone. Then, it is time to change the reinforcement zone. Of course, the arm will first seek the zone in its former position. The thing is that as far as the arm will not receive any reward, it will wander again. Then, it will find the new location. However, it takes less time to reach the new zone than the first one. This fact shows what psychologists call a capacity of generalization. We can go on and on. Each time, the arm develops a smooth and direct movement. The more it has been trained on different positions of the reinforcement, the quicker the arm finds new positions. Finally, we can also reset the reinforcement zone to a former position. Then, the arm finds this former position very quickly. So, the arm has learnt many positions of the reinforcement zone and has developed the ability to reach it by smooth movements.

Figure 4. The development of the reaching movement: from left to right, the arm first has an erratic movement; after a while, it occasionally comes through the reinforcement zone; the movement becomes smoother and smoother; finally, the arm has a direct and smooth movement. The cross indicates the shoulder of the arm. The circle is the reinforcement zone. In the 3 leftmost plots, the position of the hand is indicated. In the rightmost plot, the whole arm is sketched.

Even though the 2 dimensional arm described here may seem too simple as a system, we have also obtained identical results for a 3 dimensional arm and we are working towards a multi-armed and multi-legged animat, as well as a real robot embedding the same techniques.

5 Discussion

In this contribution, we have summarized some recent works we have done and we have drawn links between these separate results. Our point is three-fold: first, we have shown that the law of effect can result from natural selection: as the law of effect models learning during lifetime of living beings, this first point shows that the way the behavioral repertoire of a living being evolves can have been produced by natural selection along generations since the emergence of life on earth; second, we have shown that agents following the law of effect obtain good performances in a social situation as they come to cooperate; third, using Q-learning as an implementation of the law of effect (that is, Q-learning is considered as an efficient implementation of the law of effect, that is, also the product of natural selection), we have shown that a set of cooperative Q-learners used in a realistic way to simulate an arm has a dynamics which is very similar to that of the arm of a human baby acquiring the same movement. All that work has been done so that artificial agents are built using realistic and well-tested hypotheses with regards to living beings. We think that such a multi-level way of tackling the problem of acquiring new behaviors is highly interesting for two reasons: for the scientists who study living beings, this shows that complex behaviors may be explained using the law of effect, itself being the product of natural selection; for scientists aiming at creating artificial adaptive agents either in software, or in hardware, this work shows that a careful implementation of "natural" laws may give rise to complex self-organized adaptive behaviors.

This work exemplify a multi-stage selectionist process: once the law of effect have been selected and is rather widespread in a population, the cooperation between agents becomes more probable which, in turns, can yield more complex structures among adaptive components.

References

1. J.M. Baldwin. A new factor in evolution. *The american naturalist*, 30, 1896. reprinted in [7, pp. 59-80].
2. A.G. Barto. Reinforcement learning and adaptive critic methods. In D.A. White and D.A. Sofge, editors, *Handbook of intelligent control: neural, fuzzy, and adaptive approach*, pages 469–491. Van Nostrand Reinhold, 1992.
3. S. Delepoulle. *Coopération entre agents adaptatifs ; étude de la sélection des comportements sociaux, expérimentations et simulations*. PhD thesis, Université de Lille 3, URECA, Villeneuve d'Ascq, October 2000. Thèse de doctorat de Psychologie.
4. S. Delepoulle, Ph. Preux, and J-C. Darcheville. Learning as a consequence of selection, 2001. (submitted).
5. J.W. Donahoe and D.C. Palmer. *Learning and Complex Behavior*. Allyn and bacon, 1994.
6. G.E. Hinton and S.J. Nowlan. How learning can guide evolution. *Complex Systems*, 1:495–502, 1987. also reproduced in [7], chapter 25, pp. 447-454.
7. M. Mitchell and R. Belew, editors. *Adaptive Individuals In Evolving Population Models*. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley Publishing Company, 1996.
8. D. Parisi and S. Nolfi. The influence of learning on evolution. In *[7]*, pages 419–428.
9. Ph. Preux, S. Delepoulle, and J-C. Darcheville. Selection of behaviors by their consequences in the human baby, software agents, and robots. In *Proc. Computational Biology, Genome Information Systems and Technology*, March 2001.
10. J.B. Sidowski, B. Wyckoff, and L. Tabory. The influence of reinforcement and punishment in a minimal social situation. *Journal of Abnormal Social Psychology*, 52:115–119, 1956.
11. B.F. Skinner. Selection by consequences. *Science*, 213:501–514, 1981.
12. J.E.R. Staddon. *Adaptive behavior and learning*. Cambridge University Press, 1983.
13. R.S. Sutton and A.G. Barto. *Reinforcement learning: an introduction*. MIT Press, 1998.
14. E.L. Thorndike. Animal intelligence: An experimental study of the associative process in animals. *Psychology Monographs*, 2, 1898.
15. E.L. Thorndike. *Animal Intelligence: Experimental Studies*. Mac Millan, 1911.