# A Link Between Human Behaviour and Computers' Future Behaviour

Merciadri* Luca
Luca.Merciadri@student.ulg.ac.be

**Abstract.** Many studies have been written about children's development, and their behaviours during the phases they encounter before becoming adults. Thus, some clear links can be established between children's development and A.I. future trends. We here summarize the most important comparison characteristics.

**Keywords:** psychology, affective computing, human-computer interaction, emotions, reinforcement learning (RL).

## 1 Introduction

In [6], we established a very concise summary of the computer's situation, compared with humans and their emotions.

Similarly, many studies have been written about children's development ([8]), and a parallelism can be written between children's development, including their behaviours during all the phases they encounter before becoming adults, and A.I. fields.

If the computer was allowed to be more human, it would result in actions, learning, and affectivity. We detail all these characteristics in the following sections. They are also compared to the main subject: children's behaviour.

## 2 Actions

Some facts are known about "the *action*" ([8]): an action always suppose an interest, and every action answers to a need.

We would want the feelings-capable computers to perform some requested actions (such as helping autistic children, as given in [6]).

So, to perform the actions we want them to do, we (us and them) have to share the same interests, and they need to answer to their needs, which have to coincide with ours.

As Claparède showed it, a need is always the manifestation of a given desequilibrium ([3,8]). Thus, their desequilibrium has also to be the same than ours, to coincide with our need.

Briefly, to the first question

$$\text{“Who could create the } \textit{need}\text{?,”} \qquad (\clubsuit_1)$$

we would answer that *we* must create the *need*, and A.I. devices and robots should serve us, by sharing the same desequilibrium.

The serving utility of robots and A.I. devices had already been mentioned in [6]: computers are made to serve us, and should always perform this action, whatever they are asked to do. On one hand, the actions we would ask them to perform should *only* answer to *our* needs. On the other hand, all the actions they would perform would be human-oriented.

Thus, every action a system which has A.I. capabilities would do has to be performed in a view such that human is the primary interest.

## 3 Learning

As an aim of A.I. is to perform robots that *learn* by an autonomous way, different ingenious processes should be used. An example of progressive *learning* is clearly given by children.

It is known that, during babyhood and childhood, the infant learns gradually to imitate, even if there is no inherited techniques of imitation. It is the same for the sounds. This phenomenon ([8]) is commonly called "Action Socialisation," and could be useful in

* The author is thankful to Merciadri Liliane, B.Sc and M.Sc in Psychology from the University of Liège, for the "peer reviewing."

developing autonomous learning patterns in different A.I. systems, such as robots.

Its interest is significant because the children have a complex behaviour when reacting to many experiments, as shown in [8]. Studying a complex learning way can anyway have major interests in *our* perception of the potential machine's auto-learning. For example, one could be inspired by children's behaviour, and take the "inconvenients" of these children to make fastly-learning machines, heeding children's mistakes.

Evidently, we want machines to learn as fastly as they can, and we thus want to minimize their errors, if possible. The respective behaviours of children during the phases they go through engender strange (and false, until a given point) answers to some simple experiments, and undesirable effects (we call them "inconvenients"). A clear example of these undesirable effects is the fact that the most of children affirms, but never proves.

That is, machines equipped with artificial intelligence (we can clearly think to the computer notion here) would also be useful for Automated Theorem Proving (ATP), which is an important sector in A.I., as it is included in the Automated Reasoning (AR) field. The ATP would also be useful in proof verification. For example, interactive theorem provers require a human user to give hints to the system, but the theorems to prove can come from other disciplines than Mathematics: it can be used in law, engineering specifications, and computer programs.

### 3.1 With Moderate Interest

The *interest* is the extension of the needs, and thus the connection between an object and a need, because an object becomes interesting when it answers to a need ([8]).

We said earlier (at the point 2) that the *only* needs robots would have to answer to would be ours. As we do not want robots to become as autonomous as they could literally "replace us," their interest in doing things – whether we judge them interesting or not – must be systematically controlled, to prevent them from creating *their* proper needs. Anyway, using this scheme, their interests would be almost identical to ours.

### 3.2 Reinforcement Learning

One key concept of A.I. is the *reinforcement learning*. In this model, an A.I. system is conceived as to maximize a long-term *reward*[1]. This key concept is implemented using different theories, such as MDP (Markov Decision Processes). A Markov decision process is a list of four objects $(S, A, P_a(\cdot,\cdot), R_a(\cdot,\cdot))$, where

- $S$ is the state space;
- $A$ is the action space;
- $P_a(s, s') = \Pr(s_{t+1} = s' \mid s_t = s,\ a_t = a)$ is the probability that action $a$ in state $s$ at time $t$ will lead to state $s'$ at time $t + 1$;
- $R_a(s, s')$ is the (expected) immediate reward received after transition to state $s'$ from state $s$ with transition probability $P_a(s, s')$.

The aim is here to maximize some cumulative function of the rewards, typically the discounted sum over a potentially infinite horizon, thus maximizing

$$\sum_{t=0}^{+\infty} \gamma^t R_{a_t}(s_t, s_{t+1}),$$

where $\gamma$ is the discount rate (satisfying $0 < \gamma \leq 1$). Typically, $\gamma \simeq 1$. It can be compared to a rider on a bicycle. Sometimes, the tilt between the road and the rider is different than $\frac{\pi}{2}$ radians, but never rises $k\pi$, $k \in \mathbb{Z}$; thus, conceptually, the rider never falls. Progressively, he learns how to be stable on the bike, and ride better and better.

The reinforcement learning is not the same as the supervised learning, which, among other things, aims to present input/output pairs, and requires a set of questions with the right answers.

Parallely, to the interests or to the values linked to the proper activities are linked the humans' *auto-valorisation* feelings: the famous "inferiority feelings" or superiority ones (which are also felt by children).

All these successes and failures of the proper activity are recorded –in the human– in a kind of permanent value scale, the success arising the pretentions of the subject, and the failures pushing down them about future actions ([8]). This scheme is clearly linked to the relationship between the human and himself, and is also the principle of the reinforcement learning.

[1] According to [1], a *reward* is "an immediate, possibly stochastic, payoff that results from performing an action in a state".

## 4 Affectivity

As some of the upcoming robots would be able to feel, the *affectivity* plan must be considered, for the first time for *them*.

It has been proven that there exists a narrow parallelism in the affectivity's development and in the one of intellectual functions ([8]).

Thus, we cannot avoid this kind of human reactions, under penalty of losing every hope we put in emotions-capable machines, because affectivity and intellect are linked.

For example, it is always the case for Humans: even if some intelligent-seeming persons seem not to show affectivity, they cannot be *intelligent* if they do not *feel*. Briefly, they cannot be separated without causing deplorable troubles.

## 5 A Kind of Respect

As robots would continue to obey us, a clear *respect* concept is appearing here. As Bovet showed it, the respect is the source of the first moral feelings ([8]). This is not really astonishing: we spoke about computers' feelings in [6], and showed their importance in some specific domains.

As this can seem completely implausible to also respect robots or other A.I. devices, this respect must not be one-way. In fact, it suffices that the respected beings give to the person who respects them some orders, and *consignes*, to let them be felt helpful and engender too the sense of duty ([8]). Here, the "respected beings" would be us, and the "person who respects (...)" would be robots and A.I. devices. Thus, it has a clear link with reinforcement learning: thinking machines should be respected, in order to feel helpful, and to improve their productivity.

Another fact is that the language is a necessary (but not sufficient) condition to the logical operations' building ([8]). Robots should then be able to speak, and to feel heard, to continue building logical operations.

Actually, it confirms our hypothesis linked to the potential interest which could be extracted from robots and A.I. devices if they are "correctly respected." Let's resume the "ingredients" we cited.

- Firstly, giving them a *real role* in our society;
- Secondly, *avoiding* them from having *too much feelings*, except for specific uses;
- Thirdly, respecting them for the role they would be playing (*i.e.* allowing them to speak freely), and *accepting their affectivity*.

The key concept is thus *mutual* respect. There is mutual respect when individuals attribute reciprocally an equivalent personal value, and do not only valorize one of their particular actions ([8]).

It corroborates the fact that they should be treated as helpers to us. Not simple helpers, but useful machines we need to improve our productivity on some fields (as said in [6], various fields are requesting for emotions-capable machines, such as helping children).

## 6 An Equilibrium

Three points have to be kept in order to speak about *equilibrium* ([8]).

- Firstly, it is stable (but stability does not necessarily mean immobility);
- Secondly, every system can go through exterior disturbances, which tend to modify it. There is an equilibrium when these exterior disturbances are compensated by the subject's actions, oriented in compensation's sense;
- Thirdly, the equilibrium is essentially active. The more the equilibrium, the more the activity.

Speaking about equilibrium is essential, in designing A.I. devices and robots, because all these systems have to be in a certain equilibrium to satisfy our needs. According to the given definition, our A.I. future and human-like robots should then be stable, and support exterior disturbances[2], thanks to their actions, oriented in compensation's sense.

If they are stable, they will support exterior disturbances, continuing to do peaceful actions. To conceive them stable, they must respect the three rules given at point 5. It is the case if we conceive them as showed in this document.

Thus, conceiving them in a stability-vision, they will be stable, supporting exterior disturbances, and causing no troubles.

[2] By an *exterior disturbance*, we simply want to evoke facts such as humans' misbehaviours resulting in human-machine lack of respect, in interactions with emotions-capable machines.

## 7 Would They Really Be Living

Is *living*, at the beginning, every object which carries on an activity ([8]). What can we conclude, reading this definition? Strictly speaking, all the objects carrying activities are *living*. Especially, it is clear that robots and A.I. devices would be *living*.

But, again, one can feel threatened. In fact, we are only here in an *infantile animism*, defined as the trend to conceive things as if they were living, and intentionful ([8]).

Nowadays, we are finally conceiving (not only by thought, but also by engineering) many machines and gadgets that are able to carry activities. Thus, one could ask the second question

$$\text{“Are the objects we conceive really intentionful?”.} \quad (\clubsuit_2)$$

Let's take two examples. A toaster is, strictly speaking, able to carry on activity, but only when it is asked to. When it is not the case, it performs nothing, and it is thus never *living*.

The toaster's example is clear, but take now a robot, which *shows* feelings. When can we consider it as *living*? It can carry on activity, in a more or less autonomous way, but is it really intentionful? The distinction is now a little bit less clearer.

Strictly speaking, many of the robots which are conceived today have no mind. Consequently, they cannot have intentions, and are thus not intentionful.

Thenceforth, we are nowadays still not using intentionful devices. There is thus a great difference between the devices we are accustomed to use, and the future devices which could be developed. They would *now* be *living*, and that would make the *difference*.

## References


1. Dayan, Peter and Watkins, Christopher JCH, *Reinforcement Learning: A Tutorial*, Reinforcement Learning, (2001).
2. B. Givan and R. Parr, *An Introduction to Markov Decision Processes*, 2001.
3. Hameline, Daniel, *ÉDOUARD CLAPARÈDE*, International Bureau of Education, (2000).
4. Harmon, E. Mance and Harmon, S. Stephanie, *Reinforcement Learning: A Tutorial*, , (2000).
5. Kaelbling, Pack Leslie and Littman, L. Michael and Moore, W. Andrew, *Reinforcement Learning: A Survey*, Journal of Artificial Intelligence Research, (1996).
6. L. Merciadri, *Computers and Emotions.* This article simply gives a basic overview about computers and emotions, and their future roles., 2009.
7. Pfenning, Frank, *Automated Theorem Proving*, 2004. Draft of Spring 2004.
8. Piaget, Jean, *Six études de psychologie*, Folio essais, Denoel ed., 1964.
9. Wikipedia, *Markov decision process*, 2009.


The articles coming from Wikipedia have naturally been rechecked by myself. The URI related to these documents have voluntarily not been given, because these pages are often subject to numerous modifications.