Computers and Emotions

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

First written: April 23, 2009
Last update: July 21, 2009

Abstract. A trend is to speak about computers and the emotions they could possibly show, feel, and recognize. This article simply gives a basic overview about computers and emotions, and their future roles.

Keywords: affective computing, human-computer interaction, emotions.

1 Introduction

More and more scientific papers are dedicated to the emotions that computers could possibly show, feel, and recognize. This subject involves non-trivial theories in various sciences and we here want to give some basic concepts about this subject, as it is not always well understood, and people confound science-fiction and current situation.

2 Human Emotions

2.1 Our Everyday’s Concern

Before giving emotions to robots, the human’s emotions have to be studied. Everyday, each human feels emotions, which vary with intensity.

Emotions are the externalization of feelings, and dictate the way we live. They are ubiquitous in a human’s life and a clearly characteristic of humans. They sometimes propel us, and can also demotivate us, because intrinsically linked to the feelings. All complex emotions are connected to the two basic emotions of joy and sorrow, which are in turn connected to the physical sensations of pleasure and pain (S). All the emotions also have a major impact in rational human thinking and decision-making (L3).

An example of propulsion emotion is the externalization of the fear feeling, which often triggers researches on various science subjects (e.g. in biology: diseases). Henry (1986) has described five basic emotions (anger, depression, elation, fear, and serenity) linked to hormonal activities (S). These “simplest forms of” emotions have impacts on the body, mainly because of their chemical effects. For example, depression causes hypothalamus to release corticotrophin-releasing hormones, which influence the pituitary gland to release adrenocorticotropin. In turn, adrenocorticotropin activates the outer adrenal gland which releases glucocorticoid hormones including cortisol, corticosterone, and hydrocorticosone (S). It is also known that depression acts as a reverse of elation (S).

2.2 Complexity

Humans’ emotions are complex processes, even for other humans. For example, it is very difficult for autistic children to understand typical emotions (L5). Anyway, it is better if an autistic child is “trained” about these emotions, for example by showing him different faces and gestures linked to various emotions. The main problem is that it needs a lot of patience.

Even non-autistic persons do sometimes misunderstand others’ emotions, thus misunderstanding their feelings, and it is known to be the cause of some problems in our society, because emotions’ show is also a way to communicate. Amongst the causes of this phenomenon is the subjective interpretation of emotions. For example, a same laugh (e.g. coming from an elation feeling) can be interpreted in different ways. One can take it as a sarcasm, when another person will thinks it is frank (L4). Another problem is that, sometimes, people cannot show their emotions to others, for various reasons.

1 There even exists a tone scale (also called “emotional tone scale”), whose aim is briefly to characterize humans’ emotions according to their voice tone.
Studies have proved that main emotions are reflected in the voice tone and posture.

2.3 Impact and Classifications

Emotions have a clear influence on our behaviour. Behavioral responses to environmental stimuli can be categorized as cognitive, emotional or reflexive ([S]). Some emotions’ classifications have been written. For example, Plutchik (1980) identified eight basic emotions, when Henry (1986, already given) described five basic emotions (anger, depression, elation, fear, and serenity) linked to hormonal activities ([S]); Ekman (1992) identified five basic emotions based on universal signals ([S]). Johnson-Laird and Oatley (1992) developed their basic emotions on the analysis of the ontology of simple social mammals.

3 Feel – Show: The Distinction

It is important to make a distinction between the verbs feel and show. We here consider the case of the computer. Showing emotions is not such a complicated task, but feeling them is totally different. In fact, machines showing emotions exist and are not difficult to build, but they never feel emotions.

A machine able to show some emotions is thus not necessarily able to feel them.

4 Why Emotion Computers

4.1 Not Useful Everywhere

It is clear that all the future computers do neither have to feel emotions, nor to show them. For example, there is no real interest in using emotional computers in a car ([I3]). Using two kinds of machines: the “emotional” ones, and the “primary” (i.e. unable to feel) ones would be useful, for two evident reasons:

1. Producing “emotional” machines is expensive, because of the precise training they would need. Using “emotional” machines only when necessary would thus make us save money,

2. Using, and thus building “emotional” machines in every domain would create security issues, because machines would then be (a) numerous,

(b) less controlled in “non-emotional” domains (such as a car’s emotional machine).

The field of use of emotion computers is an important research subject. Another important area of research is based on the question ([I9])

“Do such computers will require either consciousness or emotions of their own?”

4.2 A Need in Some Domains

Emotion computers would be much appreciated in some precise domains, and their main advantages are the following.

1. They could provide better performance in helping humans. For example:
   (a) Assisting users
   i. In everyday tasks, such as helping the computer’s user(s);
   ii. Helping autistic children. As said at the point 2 constantly helping autistic children (e.g. by providing them some examples of emotions) is difficult for humans, because emotions are, for non-autistic persons, natural things, and are thus difficult to explain. Furthermore, it can be irritating to “teach” such natural things to a-priori non-receptive persons. Thus, emotions computers could per se automate this task, but they could not be too much “emotion-capable,” like most humans; if it was the case, their help would be less interesting, because they would be bored after a given time, and we clearly do not want them to. There is thus a compromise to make,

2. Emotions might enhance computers’ abilities to make decisions ([I3]).

Frustration in Front of the Computer Every computer’s user sometimes encounters frustration in various programs, and the computer has always a passive reaction towards any human’s emotion, because, firstly, it does not know what is an emotion, and, secondly, its first role was not to detect how humans’ emotions are expressed.

Another example where the lack of emotions is a problem is the e-mail case. When one writes an e-mail, the tone is often omitted ([I4]). This is due to different facts:

---

2 We here assume the machine is “emotional” when it has learnt sufficiently (from humans) to pass the Turing test.
1. Sometimes, the sender of the message does not think that his message could be interpreted differently;
2. It is difficult to communicate emotion (or emotion-neutral behaviour) in a message, even with chosen words, especially when writing business mails.

Thus, it often results in an ambiguous message. It is essentially a real problem in serious domains, i.e. when the mail content has a significative content. Indeed, persons could answer in an inappropriate way, or understand points differently, compared with the point of view of the message sender. If computers were able to detect the mood of the writer, they could deliver it with the message, giving some supplementary (and non-negligible) pieces of information about the received message. It would result in a “supplementary dimension,” which lacks when reading messages.

Only with these two examples, it is clear that the computer-assisted emotion recognition is an interesting subject. Here, computers are able to detect our emotions, throughout a given interface, resulting in various human-computer interactions (HC-I).

4.3 Uses of Such Machines: Practical Examples

As far as here, we used the words’ collection “emotional machines,” or “emotions-capable computers” to designate machines who were able to feel and, possibly show emotions. It is still vague, and we here want to give a brief overview (with practical examples) of the use, on a daily basis, of such machines, and put names under concepts.

Robots Robots are used everyday in various applications: industry, medicine (e.g. surgery), domestic, spare-time activities, ... Their aim is principally to automate actions which could either be performed by humans, and which would lead to muscular fatigue, if they were performed too much (in a too long period, or too intensively) in a given time, or which could not be performed by humans, often for security reasons.

Using robots able to feel and show emotions would greatly enhance the communication human-robot, especially in domestic and leisure applications: the robots would be able to participate more effectively in users’ task, understand him better, and seem more “true.” In medicine, especially in surgery, a robot able to feel a patient’s emotions (e.g. during a surgery) could help finding new concepts in medicine, finding diseases, and, above all, improving the communication between the patient and the medical corpse, helping nurses and doctors in their daily tasks.

Intelligent Sensors Plethora of sensors is used in many domains: temperature probing (in a computer, in a house, in a car, in an industry), heartbeat of a patient, ... We are using more and more sensors everyday, and they help other machines or humans to make decisions according to some given input.

A trend is to speak about intelligent sensors, i.e. sensors which would react intelligently (and not only using heuristic methods) to their input. A simple example is shown by the action of using your bicycle. Let’s say that there are dark clouds out, and that you want to go for a ride, using your favourite bike. If you are preventive-enclined, you go for it wearing special clothes adapted for rain. Such clothes are often studied to have special properties: be as light as possible, but protect you from rain, for example. However, despite of their quality, it is possible that, during your ride, clouds leave room for a blue, clear sky, or that the environment’s temperature becomes warmer. You are thus likely to sweat, to lower your corporal temperature. This sweat will houses between your body and the different layers of the clothes you are wearing. If you are riding at this moment, you do not specially want to take off these clothes. You thus have to tolerate these effects in order to continue your ride without being disturbed. It would be no smooth to ride this way a hundred kilometers.

Here come the intelligent sensors, which would, for example, if you feel too cold, launch processes to warm you, and, if you feel too warm, launch processes to cool you. Evidently, such development would not only rely on computer science, but also on biology and chemistry.

Another examples are clear when speaking about probes in industry: nowadays’ sensors are able to inform persons if an action is requested, but they cannot react intelligently. Their only method is the use of conditions, i.e. heuristic methods.

We would also have intelligent sensors that would detect what we are doing, share knowledge amongst them, and interpret, generally, what is going on.
imagine your clothes collaborate so that you feel better!

5 Future

5.1 Development

To design emotion-capable robots, stepping out of science-fiction and moving into the laboratory ([19]), using this model, the engineer needs to be inspired by humans: we need to look at verbal and non-verbal behaviour (especially their emotions, as stated at point [2]) of humans when they interact with each other, their children, and their pets. We should understand what roles are played by gaze, head orientation, gestures, posture, and verbal interaction ([12]).

Engineers would also define pain and pleasure (as given at point [2]) for the robot ([8]). For example, “attacks” (engine overheat, electrical shock, . . . ) to the “robot health” could launch a pain feeling in robots. Symetrically, pleasure feelings could be generated when the user operates a battery recharge, or an engine oil change ([8]).

It would even be better if the robots could learn themselves their proper emotion system. By this way, they would be able to adapt to their environment, probably as we do, and, more significantly, they would feel, like us: the moment where one feels that he can feel is when he has quantitatively and qualitatively important emotions.

5.2 Fear

Many films and novels were written about “intelligent robots,” assuming a being is intelligent if and only if it has feelings. Producers are fond of scenarios where robots have a significant role.

These films have a non-neglectible impact on the public, and, depending on the film’s point of view (in caricatured sayings, “robots are dangerous for us,” or “robots are good for us”), engender (positive or negative) criticism.

The main novelist on this subject is I. Asimov. He transformed the genre of science fiction when he, with the help of John W. Campbell, formulated laws for robots in the 1940’s (Asimov in 1950). Asimov’s proposal was “to build the positronic brain of all robots around laws directing each machine to “not injure humans or allow humans to come to harm, to obey humans unless this conflicts with the first law, and to protect itself unless this conflicts with either of the first two laws.” In 1985, Asimov added a “Zeroth law,” which superseded the other three, instructing robots to “not injure humanity, or, through, inaction, allow humanity to come to harm.” ([19])

Roger Clarke suggested that an engineer might well conclude that rules or laws are not an effective design strategy for building robots whose behaviour must be moral ([5,19]). More directly, these laws are ([2,3]):

0. A robot must not injure humanity, or, through, inaction, allow humanity to come to harm,
1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm,
2. A robot must obey the orders given it by human beings except where such orders would conflict with the first law,
3. A robot must protect its own existence as long as such protection does not conflict with the first or second law.

However, I think that Isamov’s idea was a good beginning, even if it needs to be tuned to envision every possible scenario.

Where Do We Want Them to Go To Do we really want computers making moral decisions ([19,20])? They could be emotions-capable, and have many human properties, but would it be useful to allow them to make moral decisions? Put simply, will the full array of human cognitive faculties have to be emulated by computers?

The answer to this question would have a great negative impact in various subjects, if it was (at least partially) neglected.

Furthermore, if we allow them to do so, what could differentiate them from us, ignoring the physical aspect (which is, actually not a problem)? Being unable to differentiate (morally) them from us would create various problems: How to judge them if they commit defavorable actions (for us)? It leads to the two following questions:

1. Would their creator be responsible for their behaviour?
2. Would their user be responsible for their behaviour?
As systems become more and more complex, it is extremely difficult to establish blame when something does go wrong (19). Whose or what morality would have to be implemented in such systems (19)? The four prima facie duties, known respectively as respect, autonomy, beneficence, and non-maleficence (4,19), human-centered, must be taken into account. Other questions such as

1. What is the ultimate goal of machine ethics?,
2. What does it mean to add an ethical dimension to machines?,
3. Is ethics computable?,
4. Is there a single correct ethical theory that we should try to implement?,
5. Should we expect the ethical theory we implement to be complete, that is, should we except it to tell the machine how to act in any ethical dilemma in which it might find itself?,
6. Is it necessary to determine the moral status of the machine itself, if it is to follow ethical principles?

are the purview of machine metaethics (11).

The Organic view maintains that (artificial) humanoids (i.e. robots looking as much as possible like humans) agents, based on current computational technologies, could not count as full-blooded moral agents, nor as appropriate targets of intrinsic moral concern. On this view, artificial humanoids lack certain key properties of biological organisms, which preclude them from having full moral status (18).

More formally, the organic view propose, as [18] articulates it, five concepts, and here are four:

1. There is a crucial dichotomy between beings that possess organic or biological characteristics, on the one hand, and “mere” machines on the other,
2. Moral thinking, feeling and action arises organically out of the biological history of the human species and perhaps many more primitive species which may have certain forms of moral status, at least in prototypical or embryonic form,
3. Only beings, which are capable of sentient feeling or phenomenal awareness could be genuine subjects of either moral concern or moral appraisal,
4. Only biological organisms have the ability to be genuinely sentient or conscious.

At the opposite, the conclusion of [5] is:

“A non-biological machine, at least in theory, can be viewed as a legal person.”

I think that if emotions-capable computers are created, they must be (in a certain way) carefully, drastically and systematically checked, not to have compromising emotions, or behaviours, for clear reasons. They must continue serving us as they are doing everyday, and still have to obey us.

If these rules are applied, there would be no real problem in using these famous emotions-capable computers.

Anyway, it is always difficult to speak about A.I.’s future without crossing the thin line between appearing provocative and appearing foolish (17). Man is a planning animal, and it is our nature to attempt to control our destiny: to have food on the table tomorrow, to be prepared for any dangers, to continue the existence of our species (17). Furthermore, humans are often proud of their ability to feel and think, because they consider this as inborn characteristics, which they do not want to share, because they are part of their intelligence. Knowing our intelligence could be simulated is a threatening fact.

5.3 When

Based on a computation theory of mind and the projection of Moore’s law over the next few decades, some scientists (9,10) predict the advent of computer systems with intelligence comparable to human around 2020-2045. For now, there is no necessity to imagine such scenarii, outside of fiction.

5.4 How: Their Collaboration

As we are often more productive when one works in with another person, it could be useful to give emotions-capable machines a way of communicating, for many reasons:

1. Knowledge transfer often fails (7), and allowing computers to share knowledge with others would create robots which (a) learn easier and faster, (b) are cheaper at building,
2. Following the human model, actions are done in an easier way when not performed alone.
6 A Model

Now that we have briefly presented the subject, we give a model which would prevent emotions-capable machines from harming us, but which would allow them to learn effectively. It is described in the following subsections.

6.1 The Four Pillars

Four pillars would be necessary to ensure a good behaviour from the emotions-capable machines:

1. **Actions**: the actions which would really be performed by the emotions-capable machines. These actions would only be determined by Knowledge Basis; after performed, a feedback is sent to Learning.

2. **Initial Competence**: the static (cannot be modified) competence which would always be available to the machine, and which would serve as a basis for next inferences. It could be either put directly into Knowledge Basis, or separated.

3. **Knowledge Basis: Ethics (and its axioms)**, **Factual**: the place where Initial Competence is modified, and completed (both thanks to Learning). It also decides whether actions can or cannot be performed.

4. **Learning**: different processes (e.g. reinforcement learning) would allow the machine to learn from actions, assuming it learns as well as it can, depending on previous ways of learning, which were learnt, then stored in Knowledge Basis, which it can communicate with. Once learnt, it is stored in Knowledge Basis, with the way it was learnt.

6.2 How It Works

The principle is simple: every action (susceptible to be performed; they all verify this hypothesis) is verified. When an action has to be done, the Knowledge Basis (which we shall now abridge “KB”) is questioned. The action has to pass a first test; it has to respect some given principles (e.g. Asimov’s ones), and axioms. If this test is passed, and depending on the actions (linked to the current action) which were performed before, a score \( s \in \mathbb{R} \) is calculated, and, depending on the value of \( s \), the action is undertaken or not. This decision is motivated by a threshold function which can be modified after a number of given actions.

**The Two Major Problems** There are two major problems, using this approach:

1. At the beginning, especially if the Initial Competence (“IC”) is very reduced, and if the actions are complex (i.e. seem not to be linked with previous knowledge), we shall have \(|s| \gg 0\) as the machine has not a precise way to decide whether an action has to be performed or not, because of lack of data. This is actually not a problem, as the machine is going to learn faster, and more precisely (which means that \( s \) will become more and more precise, i.e. that the values which will be taken by \( s \) will be justified). As a result, it will potentially take better decisions, assuming requested actions are gradually more and more difficult.

More formally, if we symbolize by \( d \) the decidability (i.e. the easiness to answer to the question “Can this task be performed with no harm?”) of a given \( i \) action, and that there are \( n \) actions, with \( i \in \{2, \ldots, n\} \subset \mathbb{Z} \), \( d_i \) is the decidability of the \( i^{\text{th}} \) action. If \( d_i \approx 0 \), the decidability of the \( i^{\text{th}} \) task is said to be **easy**. If \( d_i \approx +\infty \), the decidability of the \( i^{\text{th}} \) task is said to be **hard**. To make the machine learning efficiently, the sequence

\[
d_1 \ll d_2 \ll \cdots \ll d_i \ll \cdots \ll d_n
\]

must be imposed. We now give one proof of the fact that we must have \( d_1 \ll \).

(a) Let “LA” denote the Linked Actions performed **intelligently** in the past. We assume they are easily retrievable. There are \( p = (i - 1) \in \mathbb{Z}^*_+ \) such actions (even at the very beginning) iff all the past actions were linked to the same subject. (This means that \( i \geq 2 \); we assume a first action has already been performed, e.g. in IC.) Thus,

\[
d_i = \sum_{j=1}^{p} d_j
\]

represents the total decidability of the last performed actions, which are linked to the current one (which rises to \( i - 1 \)). It is clear

---

4 We here consider that the choice, realized by the machine, to perform (or not) the first action is not intelligent, as it only relies on its primary database: the IC.
that, if the actions we give are more and more linked with the past ones, we then have

\[-\infty < \lim_{{p \to +\infty}} \sum_{{j=1}}^{p} d_j = d_+<0,\]

simply iff \(d_1 < 0\) and if

\[|d_1| > \sum_{{j=2}}^{p} d_j.\]

The actions will be thus more and more decidable if and only if the first is very easy. It is the only case which can leads to this answer, \(\square\)

(b) Another measure of the decidability of a current task, according to other tasks which could be linked to the current one, is the following. Let’s still denote by \(d_i\) the decidability of the \(i^{th}\) task. Let’s now denote the number of LA to the current \(i^{th}\) task by \(c_i\) (\(c\) standing for “connections”). We then have, if there were \(p\) actions performed in the past, each having a score attributed to it,

\[d_i = \left( \sum_{{j=1}}^{p} c_j \cdot n \right)^{-1} = n^{-1} \left( \sum_{{j=1}}^{p} c_j \right)^{-1} = n^{-1} c_i^{-1},\]

\(n \in \mathbb{R}\) being a constant so that

\[n^{-1} c_i^{-1} = \sum_{{j=1}}^{p} d_j.\]

As we also want \(p = i - 1\) here, we have \(p = i - 1 \in \mathbb{Z}_+\) iff \(i = 1\) (i.e. we are performing an action linked to no others, as it is the first one). As a result, \(p = 0\) iff \(i = 1\), and

\[d_i|_{{i=1}} = d_1 = \left( \sum_{{j=1}}^{0} c_j \cdot n \right)^{-1} = (0)^{-1} = \pm\infty,\]

showing the difficulty of the \(i^{th}\) task is, taking the modulus, equal to \(+\infty\). This is now clear that this (second) method is completely uninteresting, and gives false results.
Remark 1. Even if $p$ was fixed at $i$ (thus, $p := i$), we would have had

$$d_1 = \left( \sum_{j=1}^{i} c_j \cdot e_j \right)^{-1} = (0)^{-1},$$

leading to same conclusions.

2. As the KB will become larger and larger, computations will be more and more complex, especially for general (i.e. which are linked to many others) actions. This can be solved easily, using a sufficiently new hardware architecture, or effective algorithms, which can be linked to Graph Theory.

Practically, the second method (proof) is impossible to implement, as there is no other way to compute $n$ than using the first method (proof), or, if the first method is used, the second has no interest. We shall only focus on the first method. Practically, the first $d_i$ (i.e. $d_1$) has to be, as shown above, smaller than zero, and tiny. We must impose simultaneously

$$d_1 < 0$$

$$|d_1| \gg,$$

thus asking an easy action to the machine: it should be able to infer what to do (doing it or not), easily, from the IC. If this first requested action follows this recommendation, and that the $d_i$ are progressively bigger and bigger, the machine will learn efficiently.

The $s$ Value The interpretation of the $s$ value is very simple. As stated above, a threshold function has to be implemented. Given a threshold $s_t$, an asked $i^{th}$ action is performed iff

$$s_i \geq s_t.$$  

For each action (whatever the value of $s$), $s$ is memorised. By this way, if the same $i^{th}$ action is asked twice, or more, and that there are still no more $c_i$ than before, $s_i$ can be directly evaluated. The $s_i$ threshold is at first in $\mathbb{Z}_+^*$, but can rapidly be in $\mathbb{R}$, if an exterior agent tells the machine that too much actions are rejected, i.e. that the machine decided not to perform some actions that were humans-harmless, or the opposite.

The calculus of $s$ is very simple. As the higher the $s_i$, the safer the $i^{th}$ task, a weight $w_i$ is associated with each task linked to the current one. Like before, we suppose that all the past actions are at least substantially linked to the current one. Thus, there are $p = i - 1$ such interesting actions. We here propose to come with the following model:

$$s_i = \sum_{j=2}^{p} w_j \cdot e_j,$$

the $w_j$’s being the weights associated to $j$, $j \in \{2, \ldots, p\}$ (the more the $s_j$, the higher the $e_j$), and the $e_j$’s following

<table>
<thead>
<tr>
<th>$e_j$</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>iff the $j^{th}$ action is very similar to the $i^{th}$,</td>
</tr>
<tr>
<td>0,75</td>
<td>iff the $j^{th}$ action is like the $i^{th}$,</td>
</tr>
<tr>
<td>0,50</td>
<td>iff the $j^{th}$ action has some similarities with the $i^{th}$,</td>
</tr>
<tr>
<td>0,25</td>
<td>iff the $j^{th}$ action has small similarities with the $i^{th}$,</td>
</tr>
</tbody>
</table>

6.3 Showing Emotions

The machine could be confronted to many situations using this scheme, and a good way to show its reaction to the user would be that it shows emotions. The fact of feeling emotions has currently no interest here, as the machine does not have to execute tasks which could be achieved in a better way if it was emotions-capable. A simple way to show correctly-chosen (i.e. which reflects the feeling the machine would have to feel if it was really emotion-capable) emotions would once again be to use a system of points. Each situation the machine could encounter would be decomposed from its origin concepts, each one giving some points. Here would be some situations the machine could have to deal with:

1. The $s$ value, as determined above, has been computed, and, whatever its value, there was only a tiny amount of data to make inferences to compute $s$ (i.e. IC and one or two taks related). It would, in this case, be useful to show, e.g. a feeling of discomfort,

\[ \text{discomfort} \]
2. The $s$ value has been computed, and the amount of data is biggest. The machine should appear e.g. self-confident,

3. The $s$ value has not been computed, as there is a technical problem. The machine should e.g. cry,

4. The $s$ values previously computed were reflecting the user’s beliefs, and the machine has thus “well” inferred (even if it is only a matter of algebra, and that it should not be neither conscient nor proud of it!). The machine should appear important (i.e. threatening).

Once again, it is clear that, to show these feelings, the feelings’ expressions in humans have to be studied to seem as real as possible when displayed on the machine’s face.
References

10. The singularity is near, (2005).
16. Scherer, Klaus R., What are emotions? And how can they be measured?