

Emotions: a Bridge Between Nature and Society?*

Rodrigo Ventura

Institute for Systems and Robotics
Instituto Superior Técnico
Av. Rovisco Pais, 1; 1049-001 Lisboa, PORTUGAL
rodrigo.ventura@isr.ist.utl.pt

Abstract

The field of Artificial Intelligence has, for a long time, neglected the role of emotions in human cognition, with few but notable exceptions. This has been motivated in part by the assumption that the emulation of human rationality by a machine is sufficient for attaining general human-level intelligence. This paper reviews neuroscientific results showing empirical evidence, consistently for over a decade, sustaining that emotion mechanisms in the brain play a fundamental role in decision making processes, as well as in cognitive regulation. Moreover, this role takes place regardless of whether the subject is aware of any emotion. These mechanisms are particularly important in social contexts. Lesions in the pathways supporting these mechanisms provoke serious impairments on social behavior. For instance, subjects with lesions in the pathways between the orbitofrontal cortex and the amygdala are no longer able to sustain a healthy social life, despite their intact intellectual capabilities. Strikingly, these patients are even able to verbally describe what would be the proper social behavior, although are unable to follow it. One important mechanism in social contexts is empathy, fundamental for proper social relations. It has been proposed that empathy is founded on mechanisms analogous to the mirror neurons.

Introduction

Although the sciences of the artificial may call for models of different nature than the ones used to understand emotional phenomena in humans (Simon 1996), we claim there are two reasons why understanding emotion mechanisms might contribute for the design of better intelligent machines. The first one concerns the relevance of biologically inspired models. The best model of intelligence, excluding well-defined, structured domains, still comes from humans. Thus, taking models of natural intelligence, with origin in psychology, neuroscience, or philosophy, is a natural approach, that has been proving fruitful since the advent of the Artificial Intelligence (AI) as a field. The second reason is founded on neuroscientific evidence sustaining that emotions and intelligence cannot be untangled, as far as human intelligence is concerned (Pessoa 2008).

*This work was supported by the Portuguese Government — FCT (ISR/IST plurianual funding).
Copyright © 2009, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

The reason why this point of view is not broadly acknowledged by the community is because, in our point of view, intelligence is often identified with rationality, often seen as opposed to emotional behavior. Moreover, the latter is usually understood as an epiphenomena interfering with the normal functional of the rational mind.

Emotional phenomena is in fact very broad in terms of their manifestations. Hudlicka (Hudlicka 2009) distinguishes four different modalities: (1) behavioral/expressive, which concern expression and are visible by other persons (e.g., facial expressions), (2) somatic/neurophysiological, involving changes in the body state (e.g., heart rate), (3) cognitive/interpretative, concerning their implications in the cognitive processes in the brain, and (4) experiential/subjective, which relates to the first-person subjective experience of emotions. In this paper we will be concerned with the cognitive/interpretative modality only, and particularly in what concerns decision-making. However, it should be made clear that these four modalities do strongly interact.

On rationality and intelligence

The common sense idea of rationality opposing emotions can be traced back to the Greeks. For instance, Plato sustained a ever lasting struggle between reason and emotion in our minds, with each one reaching for dominance over the other (Lyons 1999). This dualistic view lies behind the assumption that, if human level intelligence is sought, one should focus exclusively on rationality, factoring out the emotional. Intelligence has been understood as a synonym of reason. When the AI field was born in 1956, rational models dominated the field. For instance, mathematical logic has been a key theoretical framework for modeling many aspects of intelligence (McCarthy and Hayes 1969), even common sense (McCarthy 1958).

However, a few notable exceptions can be found in the early literature of AI. Herbert Simon has sustained, as early as in 1967, that emotion mechanisms can play a fundamental role as an interrupt system for a machine with multiple goals (Simon 1967). In this case, simple goal prioritization schemes can be applied, but under real-time constraints and survivability concerns, such schemes are inadequate. Alternative schemes, comparable to emotions mechanisms in humans, would then be necessary. In a similar line of reasoning, Sloman has sustained the need of emotional mecha-

nisms in artificial agents on his “Why robots will have emotions” paper of 1981 (Sloman and Croucher 1981).

Despite many successes accomplished by the field, being Kasparov’s defeat to Deep Blue in 1997 one popular account of that success (Hamilton and Hedberg 1997), general intelligence constitutes, still, largely an open issue (Nilsson 2005). We argue on the assumption that one important missing link in the understanding of general intelligence is the role of emotions. The factorization of mental activity into reason and emotions, which has been silently assumed in many approaches to AI, has been questioned by neuroscientific evidence. The next section discusses some of these findings.

On emotions and decision making

Damásio was among the first neuroscientists to question reason as a separated process from emotion. Although the modulatory effects that emotional phenomena induces on mental activity were well known (e.g., attention focus, memory retrieval, etc.), he sustained that emotions are an integral part of decision-making processes. Moreover, he stresses that these mechanisms are founded on the body, and thus mind and body make an indivisible whole. This contrasts with the dualistic mind-body view of Descartes, thus motivating the name of his book “Descartes’ Error” (Damásio 1994). This view is founded on his Somatic Marker Hypothesis (SMH), according to which mental imagery is associated with internal representations of body states (Damásio, Tranel, and Damásio 1991). In certain situations (e.g., stressful), the brain associates mental imagery related with a situation with the alterations of the body state representations, induced by the emotional state. The associations thus formed can be reenacted later, when the subject is experiencing a similar situation, or even when considering that situation as a possible consequence of a course of action. This reenactment occurs using the same brain mechanisms as the ones prompting the body state alterations following the emotion. This brain zone is the amygdala, and plays a central role in virtually all emotion processes in the brain.

The amygdala is responsible for the body state implications of an emotion, in the two kinds of emotions distinguished by Damásio (Damásio 1994). The first one is primary emotions, which correspond to emotions immediately following sensory stimuli. For instance, the startle response after a sudden and unexpected noise. The second one is secondary emotions, that correspond to body state alterations induced by emotionally loaded mental imagery. For instance, when considering performing an action that can bring a painful consequence. In both cases, the amygdala is responsible for altering the body state¹.

The implications of these processes in decision making comes from a set of projections from the amygdala to the prefrontal cortex, most high-level, cognitive processes are believed to occur (reasoning, planning, working memory, and so on). The meaning of these projections from the

amygdala to the prefrontal cortex was studied by Damásio. It is from studies of patients with lesions in these projections that most evidence supporting the SMH comes from. Patients with these lesions behaved otherwise normally, except when facing certain decisions. The cognitive capabilities as evaluated by I.Q. tests turned out to be within normal. Damásio describes a particular case of such a patient that, when faced with the need to schedule his next meeting with him, the patient was unable to do so in useful time. He pondered endlessly the pros and cons of each possible option. Other reported consequences are the inability to make reasonable financial investment decisions, difficulties in initiating a loving relationship. These patients usually lose their jobs, and marriages often dissolve. This suggests that the most practical, daily forms of decision making depend critically on emotional mechanisms in the brain (Damásio 1994; 1999).

For a quantitative evaluation of the SMH, Damásio and colleagues set up an experiment (the Iowa Gambling Task) exposing subjects to decision making situations under uncertainty (Bechara et al. 1997). In short, this experiment consisted in a one person game, where the patient was asked in each turn to choose a card from one of four decks, labeled from A to D. A loan of \$2000 was initially given. Decks A and B most often yielded gains of \$100, but occasionally implied a big loss of as much as \$1250, while decks C and D yielded most frequently smaller amounts, \$50, but the player was also subject to occasional smaller losses not higher than \$100. No information about this distribution of gains and losses was provided to the subjects. The decks were setup in such a way that decks A and B were disadvantageous, while C and D were advantageous, in the long run. The results were striking: normal control patients, after an initial sampling of all decks, tended to converge to choose decks C and D, while patients with the referred lesions tended to choose A and B.

According to the SMH, when a subject is considering which deck to choose, there are alterations in the representation of the body state, enacted by the amygdala. These alterations provide an important bias towards certain options. Patients with the referred lesions do not enact these modifications, and thus they are not subject to the same biases as normal controls. These modifications have in fact being physically measured by Damásio’s team. The lack of these biases in these patients is behind, according do Damásio, their insensitivity to the bigger losses implied by decks A and B. This insensitivity is referred by Damásio as “future myopia” (Damásio 1994).

In summary, it is in situations where future outcomes are more uncertain, in the sense of being hard to predict in detail, that emotions seem to play a more important role.

The importance of the amygdala in decision making has also been addressed by studies identifying two levels of decision making in the brain (Frank and Claus 2006). A first, more primitive one, concerns a reinforcement learning like mechanism, involving the basal ganglia and the neuromodulator dopamine, and a second one, involving the prefrontal cortex, namely its orbitofrontal region. The former, corresponding to an evolutionary older part of the brain, is slow

¹Damásio bears that there is a mechanism in the brain, that he calls the “as-if-body-loop”, that can short circuit this body consultation, but it does not replace the body entirely (Damásio 1994).

to adapt to changing conditions, namely when the association between rewards and situations may change dynamically (as in reward reversal). The latter one, located in the newer brain zones, provides a sophisticated adaptive function, as it is able to rapidly adapt to the environment (Rolls 2004). This allows for adapting one's behavioral strategy to the challenges of dynamic environments. One striking aspect is that this newer decision mechanism does not function independently, but rather it is strongly connected with the amygdala (Holland and Callagher 2004).

On emotions and social cognition

One prime example of a dynamically changing environment is social contexts. Here, the pairing between actions and reward expectancy may change dramatically with time. The Iterated Prisoner Dilemma (IPD) game is a paradigmatic example of such a situation. The IPD is an iterated version of the classic Prisoner Dilemma game: two players that, without being able to communicate, are asked to choose one of two possible options, to cooperate (C) or to defect (D). The consequences are such that, regardless of the option taken by one player, the other always maximizes his payoff by defecting. However, they both benefit most after mutual cooperation, than when they both defect. The IPD is an iterated version of the Prisoner Dilemma game, where each player has access to the previous turns, and the payoffs are monetary. From an individual point of view, each player maximizes his payoff by defecting. For example, if they both defect, the payoff is \$1 for each, while mutual cooperation yields \$2 to both of them. However, when one cooperates and the other defects, the former gets \$0, while the defector gains \$3.

With the use of brain imaging techniques, subjects have been scanned while playing IPD games. It was found that brain regions implicated with the SMH mechanism play an important role in cooperative behavior in this game (Rilling et al. 2002; Adolphs 2003). In particular, the mechanism of reenactment of the body state representation seems to be crucial for subjects to cooperate. Failure to do so, as can be observed in patients with certain lesions, lead subjects to defect, preferring immediate rewards, in exchange for the long term benefits of mutual cooperation. Interestingly, after the experiments, subjects reported that mutual cooperation was the most personally satisfying outcome, while defection provoked feelings of guilt.

One of the serious consequences of the lesions studied by Damásio concerned social behavior. From his studies of patients with lesions affecting emotional mechanisms, he reported that they lose the ability to make appropriate decisions under uncertainty. For instance, they showed impairment in maintaining personal trust, empathy, adequate social behavior, maintaining marriage and an healthy relationships with the offspring. But strikingly, intellectual capabilities remained intact, as they were (verbally) aware of the social rules they themselves break (Damásio 2003).

One important mechanism for social relationships is empathy. Brain imaging studies have revealed that empathy is based on changing one's internal body representation, by one replicating the feelings of the other. One study have

shown that this representation is more intense when imitating a facial expression than when observing one (Carr et al. 2003). Another study provided evidence that we understand the pain of others by instantiating it internally in our brain (Jackson, Meltzoff, and Decety 2005). The importance of this internal reenactment of feelings of others is corroborated by evidence revealing that lesions in the amygdala compromise more the perception of social emotions from faces, than simple emotions (Baron-Cohen and Tranel 2002). This reenactment of internal body states after the observation of the same states in others resonates nicely with the discovery of the mirror neurons: these neurons are active both while performing a goal-directed action, and while watching someone else performing the same action (Gallese et al. 1996).

In summary, neuroscientific evidence have supported the importance of emotional mechanisms in the brain for appropriate social behavior (Rilling, King-Casas, and Sanfey 2008). One reason for this may be the uncertain nature of the decisions involved in social contexts. It is under uncertainty the the emotional mechanisms are more relevant for decision making, as it was discussed in previous sections. Moreover, it has been argued that at the core of human ethics and moral sense are (covert) emotional processes (Damásio 2003; Koenigs et al. 2007).

Conclusions and Discussion

This paper reviewed recent evidence sustaining the importance of emotional mechanisms in human decision making. The role of these mechanisms were discussed first at the individual level, and then in social contexts. Moreover, these mechanisms were found to be particularly important in social contexts. These results come mostly from brain imaging techniques and from comparative studies with patients with certain lesions in the brain. These findings, together with many others, suggest that, as far as human intelligence is concerned, rationality cannot be uncoupled from emotion. These two processes work together, and they jointly contribute for intelligent behavior.

Another aspect is that the amygdala lies in an evolutionary old part of the brain, but has connections to a large amount of cortical areas, thus suggesting that it serves as a hub in the brain (Pessoa 2008). The newer parts of the brain, such as the prefrontal cortex, bringing to the human species sophisticated cognitive capabilities, depend on more basic structures, such as emotions and somatic representations.

From the perspective of the design of intelligent machines, we claim that inspiration from biology is a rich paradigm for advancing the field. In this line, one should abandon the idea of rationality and emotions as two separate processes, as apparent by naive introspection. Rather, it should be understood that intelligent behavior in humans is an integral result of cognitive and emotional processes.

This understanding is in fact not new, and some computational approaches to emotions and cognition integrated have been proposed in the past (Sloman 1997; Gratch 1999; Gadanho and Hallam 2002; Hudlicka 2002; Morgado and Gaspar 2004; Maçãs and Custódio 2003; Ventura and Pinto-Ferreira 1998; 2009). These efforts constitute interesting

contributions towards an unified architecture taking into account cognitive and emotions processes in an integrated way.

References

- Adolphs, R. 2003. Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience* 4:165–178.
- Baron-Cohen, R. A. S., and Tranel, D. 2002. Impaired recognition of social emotions following amygdala damage. *Journal of Cognitive Neuroscience* 14(8):1264–1274.
- Bechara, A.; Damásio, H.; Tranel, D.; and Damásio, A. R. 1997. Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–1295.
- Carr, L.; Iacoboni, M.; Dubeau, M.-C.; Mazziotta, J. C.; and Lenzi, G. L. 2003. Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America* 100(9):5497–5502.
- Damásio, A. R.; Tranel, D.; and Damásio, H. C. 1991. *Frontal Lobe Function and Dysfunction*. Oxford University Press, NY. chapter Somatic Markers and the Guidance of Behavior: Theory and Preliminary Testing, 217–229.
- Damásio, A. R. 1994. *Descartes' Error: Emotion, Reason and the Human Brain*. Picador.
- Damásio, A. R. 1999. *The Feeling of What Happens: Body and emotion in the making of consciousness*. Harcourt Brace.
- Damásio, A. R. 2003. *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Harvest Books.
- Frank, M. J., and Claus, E. D. 2006. Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review* 113(2):300–326.
- Gadano, S. C., and Hallam, J. 2002. Robot learning driven by emotions. *Adaptive Behavior* 9(1):42–64.
- Gallese, V.; Fadiga, L.; Fogassi, L.; and Rizzolatti, G. 1996. Action recognition in the premotor cortex. *Brain* 119:593–609.
- Gratch, J. 1999. Why you should buy an emotional planner. In Velásquez, J., ed., *Workshop on Emotion-Based Agent Architectures (EBAA'99)*, 53–60.
- Hamilton, C. M., and Hedberg, S. 1997. Modern masters of an ancient game. *AI magazine* 18(4):137–144. AAAI.
- Holland, P. C., and Callagher, M. 2004. Amygdala–frontal interactions and reward expectancy. *Current Opinion in Neurobiology* 14:148–155.
- Hudlicka, E. 2002. This time with feeling: Integrated model of trait and state effects on cognition and behavior. *Applied Artificial Intelligence* 16(7/8):611–641.
- Hudlicka, E. 2009. Challenges in developing computational models of emotion and consciousness. *International Journal of Machine Consciousness* 1(1):131–153.
- Jackson, P. L.; Meltzoff, A. N.; and Decety, J. 2005. How do we perceive the pain of others? a window into the neural processes involved in empathy. *NeuroImage* 24:771–779.
- Koenigs, M.; Young, L.; Adolphs, R.; Tranel, D.; Cushman, F.; Hauser, M.; and Damasio, A. 2007. Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446(7138):908–911.
- Lyons, W. 1999. *Handbook of Cognition and Emotion*. Wiley. chapter The philosophy of cognition and emotion, 21–44.
- Maças, M., and Custódio, L. 2003. Multiple emotion-based agents using an extension of DARE architecture. *Informatica* 27:185–195.
- McCarthy, J., and Hayes, P. J. 1969. Some philosophical problems from the standpoint of artificial intelligence. In Meltzer, B., and Michie, D., eds., *Machine Intelligence 4*. Edinburgh University Press. 463–502.
- McCarthy, J. 1958. Programs with common sense. In *Mechanization of Thought Processes, Proceedings of the Symposium of the National Physics Laboratory*, 77–84. London, U.K.: Her Majesty's Stationery Office.
- Morgado, L., and Gaspar, G. 2004. A generic agent model allowing a continuous characterization of emotion. In Trappl, R., ed., *Cybernetics and Systems 2004*, 651–656.
- Nilsson, N. J. 2005. Human-level artificial intelligence? be serious! *AI magazine* 26(4):68–75.
- Pessoa, L. 2008. On the relationship between emotion and cognition. *Nature Reviews Neuroscience* 9(2):148–158.
- Rilling, J. K.; Gutman, D. A.; Zeh, T. R.; Pagnoni, G.; Berns, G. S.; and Kilts, C. D. 2002. A neural basis for social cooperation. *Neuron* 35(2):395–405.
- Rilling, J. K.; King-Casas, B.; and Sanfey, A. G. 2008. The neurobiology of social decision-making. *Current Opinion in Neurobiology* 18:159–165.
- Rolls, E. T. 2004. Convergence of sensory systems in the orbitofrontal cortex in primates and brain design for emotion. *The Anatomical Record Part A* 281A(1):1212–1225.
- Simon, H. A. 1967. Motivational and emotional controls of cognition. *Psychological Review* 74(1):29–39.
- Simon, H. A. 1996. *The Sciences of the Artificial*. MIT Press, third edition edition.
- Slooman, A., and Croucher, M. 1981. Why robots will have emotions. In Drinan, A., ed., *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI-81)*. Menlo Park, California: The AAAI Press.
- Slooman, A. 1997. *Foundations of Rational Agency*. Kluwer Academic Publishers. chapter What sort of architecture is required for a human-like agent?
- Ventura, R., and Pinto-Ferreira, C. 1998. Emotion-based agents. In *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98)*, 1204. Cambridge/Menlo Park: AAAI Press and The MIT Press.
- Ventura, R., and Pinto-Ferreira, C. 2009. Responding efficiently to relevant stimuli using an emotion-based agent architecture. *Neurocomputing* 72(13–15):2923–2930.