Published in final edited form as:

Trends Cogn Sci. 2017 November ; 21(11): 817–819. doi:10.1016/j.tics.2017.06.010.

Do Intelligent Robots Need Emotion?

Luiz Pessoa[1,*]

[1]Department of Psychology and Maryland Neuroimaging Center, University of Maryland, College Park, MD, USA

Abstract

What is the place of emotion in intelligent robots? Researchers have advocated the inclusion of some emotion-related components in the information-processing architecture of autonomous agents. It is argued here that emotion needs to be merged with all aspects of the architecture: cognitive–emotional integration should be a key design principle.

Emotional Robots?

In an episode of the HBO Westworld series, a human is interrogating Dolores, a central character of the show. Her answers are charged with emotion, they convey a deep feeling of confusion and anxiety. The examiner then commands: ‘cognition only’; Dolores continues without any trace of emotion. Dolores is a ‘host’ (how the humanoid robots are called) at Westworld, a theme park created for the entertainment of humans who are free to kill or have sex with them.

In building an intelligent robot it might seem best to focus on its cognitive capacities, perhaps disregarding emotion entirely, or including only as much emotion as necessary. If emotion is included, to avoid ‘emotional interference’, the design would allow the cognitive module to downregulate emotion as desired (Figure 1A).

It is argued here, instead, that cognition and emotion need to be intertwined in the general information-processing architecture (Figure 1B). The contention is that, for the types of intelligent behaviors frequently described as cognitive (e.g., attention, problem solving, planning), the integration of emotion and cognition is necessary. The proposal is based on a growing body of knowledge from brain and behavioral sciences [1–4].

Knowledge about anatomy and physiology indicates that emotion and cognition are closely intertwined [5]. Anatomical and functional studies reveal that brain regions are massively interconnected [6,7]. In particular, the brain basis of emotion has been suggested to involve large-scale cortical–subcortical networks that are distributed and sensitive to bodily signals [5]. The high degree of signal distribution and integration in the brain provides a nexus for the intermixing of information related to perception, cognition, emotion, motivation, and action. Importantly, the functional architecture consists of multiple overlapping networks that are highly dynamic and context-sensitive. Furthermore, adopting a large-scale network

*Correspondence: pessoa@umd.edu (L. Pessoa).

perspective to brain organization [5–7] helps to clarify why some brain structures, such as the amygdala, are thought to be important for emotion: they are important hubs of large-scale connectivity systems. It also illuminates why the impact of emotion is so wide-ranging – it is not possible to impact emotion without affecting perception and cognition.

Perceptual and Executive Competition

Objects in the environment compete for limited perceptual processing capacity and control of behavior. Because processing capacity for vision is limited, selective attention to one part of the visual field comes at the cost of neglecting other parts. Thus, a popular notion is that there is competition for resources in visual cortex.

One way in which competition can be understood is via the concept of a priority map [8] which contains representations of spatial locations that are behaviorally important. Priority maps have been identified in the multiple brain regions, including frontal and parietal cortex, as well as in superior colliculus and the pulvinar nucleus in the thalamus. Traditionally, ‘bottom-up’ factors (such as stimulus salience) and ‘top-down’ factors (such as goal relevance) were emphasized as the major inputs to determine priority. Newer evidence shows that we must add affective significance [9] (e.g., based on the pairing of a stimulus with aversive events) and motivational significance [10] (e.g., based on the link between actions and rewards) to these inputs. In the brain, multiple mechanisms embed affective and motivational significance, as well as stimulus- and goal-related factors, into perception [3]. In particular, perceptual competition involves interactions between regions important for attention (in frontal and parietal cortices, for example) and those that are particularly tuned to the evaluation of affective and motivational value.

Executive control refers to operations involved in maintaining and updating information, monitoring conflict and/or errors, resisting distracting information, inhibiting prepotent responses, and shifting mental sets. A useful way to conceptualize executive control is in terms of a set of mechanisms needed for functions such as inhibition, updating, and shifting [11]. These mechanisms are not independent, however, because when a given mechanism is necessitated it will not be available to other operations and interference will ensue, possibly compromising performance. In other words, we can conceptualize this interdependence as a form of executive competition.

Dealing with an emotional stimulus or situation requires the types of behavioral adjustments that characterize executive function. For example, updating might be needed to refresh the contents of working memory, shifting might be recruited to switch the current task set, and inhibition might be invoked to cancel previously planned actions. In this manner, executive functions are coordinated in the service of emotional processing and, if temporarily unavailable to additional task requirements, performance could be compromised – and the stronger the emotional manipulation, the stronger the interference. A simple laboratory example illustrates these ideas. Suppose a participant is performing an effortful task and a change of background color signals that she will receive a mild shock sometime in the next 30 seconds. The participant may update the contents of working memory to include the ‘shock possible’ information. In addition, the participant may shift processing between the

execution of the cognitive task and ‘monitoring for shock’ every few seconds. Now, if another cue stimulus indicated that shock would be delivered in the next second, the participant might temporarily inhibit responding to the task to prepare for the shock. In other words, dealing with the emotional situation necessitates the same types of executive functions considered to be the hallmark of cognition. Importantly, the intensity of the emotional stimulus (or context) determines if it will improve or hinder behavioral performance. Moderate intensity, in particular, allows processes to be devoted to the situation at hand, improving performance. However, if the intensity is sufficiently high, processes may be temporarily unavailable to handle some of the aspects of the situation adequately (such as successfully executing the effortful task above).

Cognitive–Emotional Architecture

In the past two decades a steady stream of researchers have advocated the inclusion of emotion-related components in the general information-processing architecture of autonomous agents ([12–14], see contributions in [15]). One type of argument is that emotion components are necessary to instill urgency to action and decisions. Others have advocated emotion components to aid understanding emotion in humans, or to generate human-like expressions [13]. In this literature, including affect is frequently associated with the addition of an emotion module that can influence some of the components of the architecture.

The framework advanced here goes beyond these approaches and proposes that emotion (and motivation) need to be integrated with all aspects of the architecture. In particular, emotion-related mechanisms influence processing beyond the modulatory aspects of ‘moods’ linked to internal states (hunger, sex-drive, etc.). Emotion can be thought of as a set of valuating mechanisms that help to organize behavior, for instance by helping take into account both the costs and benefits linked to percepts and actions. At a general level, it can be viewed as a biasing mechanism, much like the ‘cognitive’ function of attention. However, such conceptualization is still overly simplistic because emotion does not amount to merely providing an extra boost to a specific sensory input, potential plan, or action. When the brain is conceptualized as a complex system of highly interacting networks of regions, we see that emotion is interlocked with perception, cognition, motivation, and action. Whereas we can refer to particular behaviors as ‘emotional’ or ‘cognitive’, this is only a language short-cut. Thus, the idea of a biasing mechanism is too limited. From the perspective of designing intelligent agents, all components of the architecture should be influenced by emotional and motivational variables (and vice versa). Thus, the architecture should be strongly non-modular.

The Dolores Test

Let us consider the humanoid robots of Westworld again. In the scene described, the central humanoid character was asked to consider and describe her situation by using cognition only. Could a human ‘lose all emotion’ and describe complex events without a trace of affect, as Dolores did? The present framework indicates that the answer is ‘no’. Unlike

Dolores, humans cannot lose all emotion and simply proceed cognitively. The brain is organized such that the mind does not behave in this manner.

But Dolores is not human, she is a robot. could a sophisticated artificial intelligence be built with separate cognitive and emotional modules? I contend that such a humanoid would be a far cry from the complex hosts of the series, whose behaviors are only possible when emotion and cognition are intertwined. Could this prediction be tested? It is possible to consider an extended version of the Turing test. Call it the Dolores test: by observing behaviors as humanoids interact with both other humanoids and humans, humans and humanoids would be confused with each other only if the latter were built according to a sufficiently integrated cognitive–emotional architecture (Figure 2).

To conclude, the central argument described here is not that emotion is needed – the answer is ‘yes’ – but that emotion and motivation need to be integrated with all information-processing components. This implies that cognitive–emotional integration needs to be a principle of the architecture. In particular, emotion is not an ‘add on’ that endows a robot with ‘feelings’, allowing it, for instance, to report or express its internal state. It allows the significance of percepts, plans, and actions to be an integral part of all its computations. Future research needs to integrate emotion and cognition if intelligent, autonomous robots are to be built.

Acknowledgments

The author is grateful to the National Institute of Mental Health for research support (R01MH071589) and Christian Meyer for assistance with figures and References

References

1. Phelps EA (2006) Emotion and cognition: insights from studies of the human amygdala. Annu. Rev. Psychol 57, 27–53 [PubMed: 16318588]
2. Salzman CD and Fusi S (2010) Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. Annu. Rev. Neurosci 33, 173–202 [PubMed: 20331363]
3. Pessoa L (2013) The Cognitive–Emotional Brain: From Interactions to Integration, MIT Press
4. Inzlicht M et al. (2015) Emotional foundations of cognitive control. Trends Cogn. Sci 19, 126–132 [PubMed: 25659515]
5. Pessoa L (2017) A network model of the emotional brain. Trends Cogn. Sci 21, 357–371 [PubMed: 28363681]
6. Modha DS and Singh R (2010) Network architecture of the long-distance pathways in the macaque brain. Proc. Natl. Acad. Sci. U. S. A 107, 13485–13490 [PubMed: 20628011]
7. Markov NT et al. (2013) Cortical high-density counter-stream architectures. Science 342, 1238406 [PubMed: 24179228]
8. Itti L and Koch C (2001) Computational modelling of visual attention. Nat. Rev. Neurosci 2, 194–203 [PubMed: 11256080]
9. Vuilleumier P (2005) How brains beware: neural mechanisms of emotional attention. Trends Cogn. Sci 9, 585–594 [PubMed: 16289871]
10. Chelazzi L et al. (2013) Rewards teach visual selective attention. Vis. Res 85, 58–72 [PubMed: 23262054]
11. Miyake A et al. (2000) The unity and diversity of executive functions and their contributions to complex ‘frontal lobe’ tasks: a latent variable analysis. Cogn. Psychol 41, 49–100 [PubMed: 10945922]

12. Maes P (1990) Designing Autonomous agents: Theory and Practice from Biology to Engineering and Back, MIT Press
13. Breazeal C (2002) Designing Sociable Robots, MIT Press
14. Ziemke T and Lowe R (2009) On the role of emotion in embodied cognitive architectures: from organisms to robots. Cogn. Comp 1, 104–117
15. Fellous J-M and Arbib MA, eds (2005) Who Needs Emotions? The Brain Meets the Robot, Oxford University Press

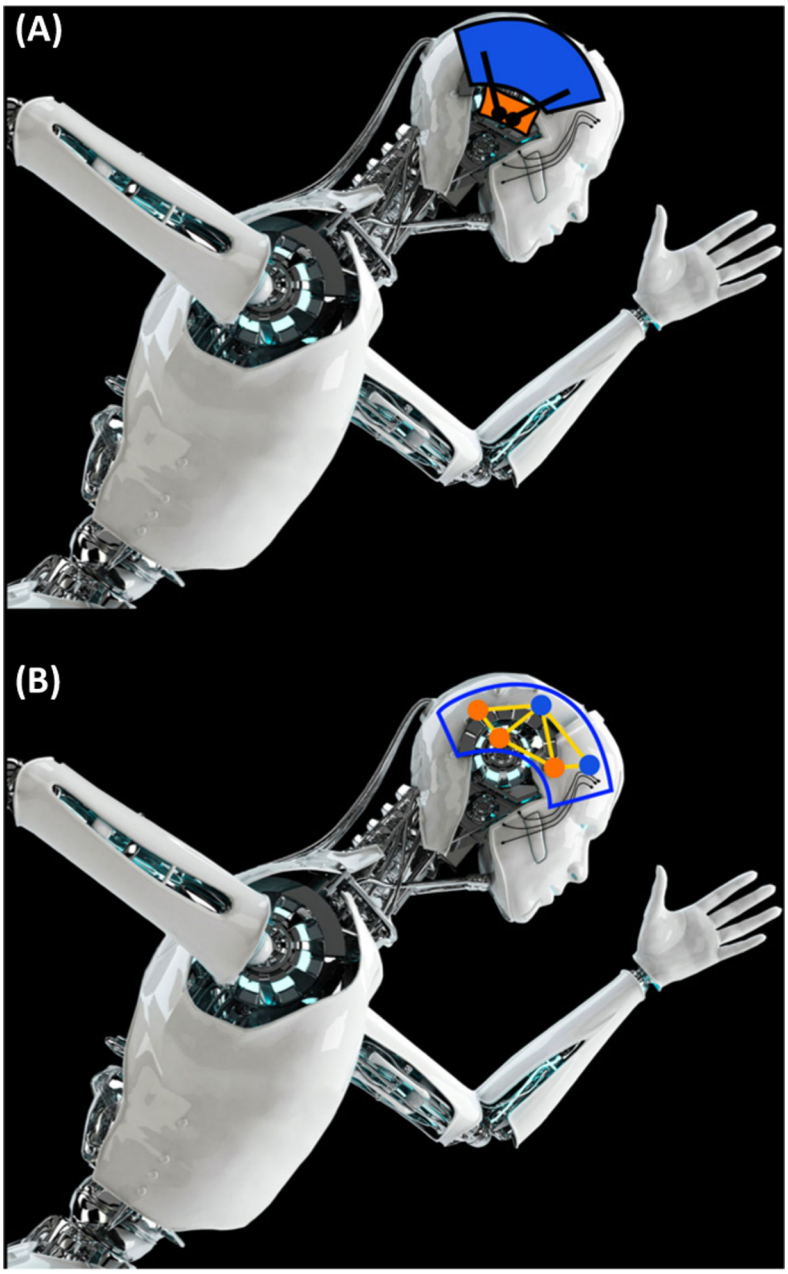

Figure 1. Information-Processing Architecture of Intelligent Robots.

(A) Modular architecture with separate cognitive (blue) and emotional (orange) components. The cognitive module also inhibits the emotion module when necessary. (B) Non-modular architecture with integrated components, which is suggested to be necessary for intelligent robots.

Trends in Cognitive Sciences

Figure 2. The Dolores Test.

The test captures the idea that cognition and emotion need to be integrated for intelligent behaviors: by observing humanoids interacting with both other humanoids and humans, humans and humanoids would be confused with each only if the humanoid information-processing architecture is sufficiently integrated. In this depiction of a scene from Westworld, Dolores interacts with two humans (one tied to the chair, another standing to her right), and is watched by two other humanoid robots.