Biologically Inspired Cognitive Architectures (2013) **6**, 109–125

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bica

RESEARCH ARTICLE

Emotional biologically inspired cognitive architecture

Alexei V. Samsonovich

Krasnow Institute for Advanced Study, George Mason University, 4400 University Drive, MS 2A1, Fairfax, VA 22030-4444, USA

Received 4 June ; received in revised form 29 July 2013; accepted 29 July 2013

KEYWORDS
Emotional cognition;
Human-level artificial intelligence;
Cognitive architectures;
Social emotions;
Affective computing;
Human–computer interface

Abstract Human-like artificial emotional intelligence is vital for integration of future robots into the human society. This work introduces a general framework for representation and processing of emotional contents in a cognitive architecture, called ''emotional biologically inspired cognitive architecture'' (eBICA). Unlike in previous attempts, in this framework emotional elements are added virtually to all cognitive representations and processes by modifying the main building blocks of the prototype architectures. The key elements are appraisals associated as attributes with schemas and mental states, moral schemas that control patterns of appraisals and represent social emotions, and semantic spaces that give values to appraisals. Proposed principles are tested in an experiment involving human subjects and virtual agents, based on a simple paradigm in imaginary virtual world. It is shown that with moral schemas, but probably not without them, eBICA can account for human behavior in the selected paradigm. The model sheds light on clustering of social emotions and allows for their elegant mathematical description. The new framework will be suitable for implementation of believable emotional intelligence in artifacts, necessary for emotionally informed behavior, collaboration of virtual partners with humans, and self-regulated learning of virtual agents.

Introduction

It was pointed recently (Samsonovich, 2012c) that the bottleneck of the long-awaited big breakthrough in artificial intelligence could be in human acceptance of virtual partners as potentially equal minds. This step would require future intelligent agents and co-robots to be able to develop mutual understanding and personal relationships with human partners at the human level (Buchsbaum, Blumberg, & Breazeal, 2004; Parisi & Petrosino, 2010). From this perspective, believable, human-level emotionally-intelligent virtual agents–partners can be expected to become a goal of research in artificial intelligence for the nearest future. It is argued here that this practical goal, although very ambitious, is achievable in the near future, at least in a limited set of domains and paradigms, and it can be achieved based on a cognitive architecture inspired by the human

E-mail address: asamsono@gmu.edu

2212-683X/$ - see front matter
http://dx.doi.org/10.1016/j.bica.2013.07.009

mind: a simple computational model capturing the essence of human emotional intelligence.

The scientific challenge motivating this work is to reconcile and unify disparate, mostly controversial theoretical views of the cognitive aspect of emotions within one universal parsimonious framework. Many limited attacks on this challenge were undertaken in the past; most recently, including works published in this journal (e.g., Sellers, 2013; Treur, 2013). Likewise, being guided by the ambitious overarching goal, the present work is limited to one small step: development and preliminary assessment of selected aspects of a parsimonious general framework of the above kind, formulated in the form of a biologically inspired cognitive architecture, or BICA.

The notion of a cognitive architecture (e.g., Gray, 2007; Langley, Laird, & Rogers, 2009) is understood here more as a framework specifying a set of principles, templates and constraints for designing intelligent agents rather than an implementation of a big system, or a programming language, like in the case of Soar (Laird, 2012; Laird, Newell, & Rosenbloom, 1987) and ACT-R (Anderson & Lebiere, 1998). This framework is supposed to be general and not limited to any specific kind of environment, embedding, cognitive task or paradigm, while its implementations may be relatively small, case-specific, and limited in many ways. A motivation for this approach taken here is the desire to learn step by step the main principles that will eventually allow us to create a ''cognitive embryo'': a ''critical mass'' of a human-level learner and an equivalent of the human mind, that will be initially demonstrated in limited settings. In the case that this will happen, emotional intelligent capabilities arguably can be expected to become vital for success. The focus of this work is therefore on principles enabling human-like emotional cognition in an agent, and on their assessment in a simplistic paradigm.

In the literature, ''emotional cognition'' may refer to cognitive processes affected by emotions, or involving understanding and awareness of emotions in self and others. Unfortunately, modern attempts to add emotional intelligent capabilities to cognitive architectures are limited both functionally (e.g., in Soar the role of emotion is limited to reward generation in reinforcement learning: Laird, 2012) and architecturally. In many cases, e.g., in the aforementioned example, in recent extensions of ACT-R (Dancy, Ritter, & Berry, 2012), and in related architectures (Marsella & Gratch, 2009), an emotion module is added as an auxiliary appendix to the existing architecture, and simulated phenomena are limited to global affective or physiological biases of the system as a whole. Previous modeling attempts to describe development of social emotional relationships among agents are limited to non-BICA approaches. It appears that a new approach in the field of BICA is required to overcome existing limitations.

The framework presented here is free of the above limitations and is an elegant way to introduce emotional elements at the core of virtually all basic cognitive process in BICA. Specifically, this framework (i) makes description and processing of emotions in a cognitive system local, (ii) makes certain known clustering of emotions natural, and (iii) adds emotional elements as intrinsic components virtually to every cognitive representation. This latter feature is arguably inspired by the human cognition. Indeed, human cognition and emotions are intimately mixed from birth and develop inseparably (Phelps, 2006).

Background on the state of the art and the problem

Recent years presented us with an explosion of research on computational models of emotions and affective computing (Hudlicka, 2011; Picard, 1997). Yet, the only existing consensus is limited to the basic affective space model that has been around for a while (Heise, 2007; Osgood, Suci, & Tannenbaum, 1957; Plutchik, 1980; Russell, 1980) and is known by different names, as briefly overviewed below. At the same time, a complete theoretical and computational understanding of the cognitive aspect of emotions is missing. In general, the notion of subjective emotional feeling is problematic in modern science: problems trace back to the general problems associated with the notion of consciousness (Chalmers, 1996). At the same time, there is no need to address the ''hard problem'' before describing cognitive-psychological and functional aspects of emotions mathematically and using this description as a basis to replicate the same observable features in artifacts.

While most modern studies of emotions do not extend beyond phenomenology, a number of views and theories were proposed that attempt to relate emotions to first principles and/or to experimental data in neurophysiology, psychology, psychiatry, sociology, theory of evolution, theory of information, control theory, and beyond. Neurophysiologically, emotional reactions and values are supported by a distributed network of brain structures, including amygdala, nucleus accumbens, anterior cingulate, paracingulate and orbitofrontal cortices, the striatum, hypothalamus, ventral tegmental area, the insula, and are mediated by major neurotransmitters, including dopamine, serotonin, acetylcholine and norepinephrine. E.g., dopamine release in nucleus accumbens results in a feeling of pleasure and is responsible for the development of emotional memories, e.g., leading to drug addiction (Kringelbach, 2009; Olds, 1956; Wise, 1978). Neurophysiological constraints like these cannot be ignored in construction of models of emotional intelligence. They help us to understand the basis of emotion; yet, they alone are not sufficient for understanding emotional intelligence. Here we focus on its cognitive-psychological rather than physiological aspect.

Psychological and computational models of emotions attempt to reduce a large variety of emotion-related concepts such as affects, appraisals, feelings, moods, traits, attitudes, intentions, motivations, and preferences to a few universal constructs (the term ''emotion'' is used here as a generalizing reference to the above factors and phenomena). Main kinds of approaches in phenomenological description of emotions (Hudlicka, 2011) include (i) taxonomies, or discrete models; (ii) dimensional models, examples of which are the semantic differential model (Osgood et al., 1957) with its variations known by different names, e.g., EPA: evaluation-potency-arousal, PAD: pleasure-arousal-domainance, Circumplex (Russell, 1980), etc.; and (iii) cognitive component models, or appraisal models, the most