Biologically Inspired Cognitive Architectures (2014) 10, 30–41

Available at www.sciencedirect.com

ScienceDirect

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

Affective facial expression processing via simulation: A probabilistic model

Jonathan Vitale [a,*], Mary-Anne Williams [a], Benjamin Johnston [a], Giuseppe Boccignone [b]

[a] *QCIS Centre — University of Technology Sydney, Ultimo, NSW 2007, Australia*
[b] *Dipartimento di Informatica — Universitá degli Studi di Milano, Milano 20135, Italy*

Received 7 November 2014

KEYWORDS
Simulation Theory;
Mirror Neurons;
Facial expression;
Latent space;
Probabilistic model

Abstract
Understanding the mental state of other people is an important skill for intelligent agents and robots to operate within social environments. However, the mental processes involved in 'mind-reading' are complex. One explanation of such processes is Simulation Theory—it is supported by a large body of neuropsychological research. Yet, determining the best computational model or theory to use in simulation-style emotion detection, is far from being understood. In this work, we use Simulation Theory and neuroscience findings on Mirror-Neuron Systems as the basis for a novel computational model, as a way to handle affective facial expressions. The model is based on a probabilistic mapping of observations from multiple identities onto a single fixed identity ('internal transcoding of external stimuli'), and then onto a latent space ('phenomenological response'). Together with the proposed architecture we present some promising preliminary results.

Introduction

In this paper we propose a probabilistic computational theory for the *detection* of emotion states based on facial expressions. This is considered a circumscribed but important mindreading task (Goldman & Sripada, 2005), either from a neuropsychological theoretical perspective and for more application oriented areas such as social robotics and social signal processing.

From a general standpoint, 'mind-reading' is the process of inferring the mental state of other people based on their facial expressions or their overt/observable behavior. It plays an important role in social interaction, empathy and effective communication. If intelligent agents and robots

* Corresponding author.
E-mail addresses: jonathan.vitale@student.uts.edu.au (J. Vitale), Mary-Anne@TheMagicLab.org (M.-A. Williams), benjamin.johnston@uts.edu.au (B. Johnston), giuseppe.boccignone@unimi.it (G. Boccignone).

http://dx.doi.org/10.1016/j.bica.2014.11.005
2212-683X/

have to interact with us naturally and in social situations, then it will be important for them to be able to 'mind-read' people.

A large body of neuropsychological research supports Simulation Theory (ST) as a plausible account for mindreading (see Goldman & Sripada, 2005, for an in-depth review). According to ST, an observer arrives at a mental attribution by simulating, in his/her own mind and body, the same state as the target. However, while the neuropsychological account of Simulation Theory is compelling, the critical question of which computational model or theory[1] might be used in simulation-style emotion detection remains poorly understood and largely unexplored (Goldman & Sripada, 2005; Pantic & Bartlett, 2007; Zeng, Pantic, Roisman, & Huang, 2009). The aim of this work is to take a step to bridge this gap by building a biologically-inspired computational account of mind-reading that is suitable for artificial agents.

In particular, we seek to apply Simulation Theory to the problem of mapping an observed overt behavior (in this work, limited to facial expressions) to a phenomenological internal latent space of the mind-reader.[2] In this study we will not consider the attribution of a specific mental state given such internal representation, as it is not the target of this paper.

Interestingly enough, computational approaches to affective state recognition from facial expression, focus on emotion labeling via direct inference of affect from facial expression (Pantic, Valstar, Rademaker, & Maat, 2005; Zeng et al., 2009). Such approaches mostly rely on Theory—Theory (TT) based accounts. In the TT account the mindreader selects a mental state for attribution to a target based on inference from other information about the target (e.g. concerning relationships or transitions between psychological states and/or behavior of the target). The rationale for such assumption is prima facie evident: postulating theory-based mechanism only requires to link facial configurations with emotion names.

However, this short-term computational advantage, which avoids a generative, simulation-based step, is, in our view, apparent. As Goldman and Sripada put it (Goldman & Sripada, 2005, p. 208): ''Simulation might be (somewhat) complex from a functional perspective, but it might be simpler from an evolutionary perspective. Simulation relies upon running the same emotional apparatus (possibly in reverse) that is already used to generate or experience the emotion. As a consequence, simulation routines do not require an organism to be outfitted with entirely new processes in order to confer an ability to recognize emotions in others''. Clearly, this parsimonious stance offers some long term advantages, for instance when coming to the challenge of integrating multiple modalities of social signaling (Zeng et al., 2009).

The remainder of the paper is organized as follows. In Section ''Background, motivations and main contributions of the proposed approach'', we provide background and motivations for a simulationist account of mind-reading. In Section ''The model'' we formalize a novel computational theory (Marr, 1982) shaped in the form of a probabilistic model. In Section ''Putting theory into work: results'' we provide one possible computational implementation of the proposed model and some preliminary experimental results. We conclude with a summary in Section ''Conclusion and future works''.

[1] The ''what'' level of explanation, in the sense of Marr (1982).

[2] In this context proposed as the internal response of the subject given a particular stimulus.

Background, motivations and main contributions of the proposed approach

The experience of emotion occurs when a complex state of the organism is accompanied by variable degrees of awareness, variously indicated as 'appraisal'. Two levels of emotion appraisal can be distinguished (Lambie & Marcel, 2002): a first-order phenomenological state and a conscious second-order awareness. Both states can be either self-directed (first-person perspective) or world-directed (third person perspective). The content of the first-order phenomenological state is physical and visceral, centered on one's body state and related neural underpinnings, that will be generically referred in the following discussion as **X**. By contrast, the content of second-order conscious awareness can be either propositional or non-propositional, which we will denote **M**. In this work, we will be concerned with first-order emotion experience, thus relying on the **X** representation. To make this intention clearer, in the rest of this paper we will use the term *detection* to describe the process of deriving a first-order phenomenological state **X** given an internal representation of an observed face expression, and *recognition* that of determining a second-order labeling **M**.

Now, we focus on the simulationist account of emotion *detection*, which can be summarized as follows:

1. In a certain situation, a subject (the *actor*) experiences an internal state $\mathbf{X}_{act}$. The internal state may be either triggered by an external event and/or mentally induced (e.g. a particular memory).
2. The internal state elicits a corresponding behavior $\mathbf{Y}_{act}$ (e.g. facial expression, gesture, heart beat, ...).
3. A second subject, (the *observer*) selects a mental state $\mathbf{M}_{obs}$ for attribution after enacting within himself the internal state $\mathbf{X}_{obs}$ in question by using as evidence the actor's behavior observable by $\mathbf{Y}_{act}$ (Goldman & Sripada, 2005).

This account is illustrated in Fig. 1.

There are several ways that this account might be translated into a computational theory. Goldman and Sripada (2005) have devised four that have substantial plausibility and are consistent with neuropsychological evidence:

1. Generate-and-test models;
2. Reverse simulation models;
3. Variants of the reverse simulation model that employ an *as if* loop;
4. Unmediated resonance models.

Generate-and-test models

Generate-and-test models assume that the observer starts by hypothesizing a certain internal state of the actor $\mathbf{X}_{act}$