ARTICLE IN PRESS

Biologically Inspired Cognitive Architectures xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

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

Research article

An android architecture for bio-inspired honest signalling in Human-Humanoid Interaction

Rosario Sorbello[a,*], Salvatore Tramonte[a], Carmelo Calí[b], Marcello Giardina[a], Shuichi Nishio[c], Hiroshi Ishiguro[c,d], Antonio Chella[a,e]

[a] RoboticsLab, Department of Industrial and Digital Innovation (DIID), Universitá degli Studi di Palermo, Palermo, Italy
[b] Dipartimento di Scienze Umanistiche, Universitá degli Studi di Palermo, Palermo, Italy
[c] Intelligent Robotics Laboratory, Graduate School of Engineering Science, Osaka University, Osaka, Japan
[d] Hiroshi Ishiguro Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan
[e] ICAR-CNR, Palermo, Italy

ARTICLE INFO

Keywords:
Honest signals
Geminoid robot
Social robotics
Human-Humanoid Interaction

ABSTRACT

This paper outlines an augmented robotic architecture to study the conditions of successful Human-Humanoid Interaction (HHI). The architecture is designed as a testable model generator for interaction centred on the ability to emit, display and detect honest signals. First we overview the biological theory in which the concept of honest signals has been put forward in order to assess its explanatory power. We reconstruct the application of the concept of honest signalling in accounting for interaction in strategic contexts and in laying bare the foundation for an automated social metrics. We describe the modules of the architecture, which is intended to implement the concept of honest signalling in connection with a refinement provided by delivering the sense of co-presence in a shared environment. Finally, an analysis of Honest Signals, in term of body postures, exhibited by participants during the preliminary experiment with the Geminoid Hi-1 is provided.

1. Introduction

Robots are going to be integrated into everyday life for cooperative, welfare and education aims due to technological innovation. Accordingly the interdisciplinary research into social robotics, Human-Robot (HRI) and Human-Humanoid Interaction (HHI) has been devoted to discovering under which conditions such integration may be successful (Kanda & Ishiguro, 2012; Lin, Abney, & Bekey, 2011; Mohammad & Nishida, 2015). This research spans in fact a wide field of features or ability that are candidates for playing a functional role in those conditions: the outward look of the robot, the implementation of cognitive and affective capacities, the display of behavioural clues that in ordinary experience display cognitive abilities (Adam, Johal, Pellier, Fiorino, & Pesty, 2016; Breazeal, 2003; Komatsu & Yamada, 2007; Sorbello et al., 2014; Walters, Koay, Syrdal, Dautenhahn, & Te Boekhorst, 2009). The variety of research lines is extended also to the designing principles that are likely to allow robots embodying the social intelligence, that is the capacities and abilities required to understand, predict and cope with other agents behaviour (Dautenhahn, 2007). One strategy is to select the set of skills, heuristics, routines, cognitive modules, which have been developed by humans and animals to solve the problems that arise living in groups whose members are tied by social bonds. That set is used to model the requirements robots have to meet to interact with humans in a common environment. An alternative option is claiming that robots, in particular humanoids, learn the required abilities by means of scaffolding (Brooks, Breazeal, Marjanović, Scassellati, & Williamson, 1999). As parents shape and guide infants acquisition of behavioural abilities and rules, so human subjects act as a scaffold that foster the required abilities in robots endowed with a motivational system as the interaction goes along (Gu & Hu, 2004). Another strategy is to build robots that are able to undergo a process akin to epigenetic development of human individuals through which they acquire intentionality, empathy and mind-reading (Kozima & Yano, 2001). In this paper we focus on the research into the minimal conditions that are reasonably the core of as successful and natural-like an interaction as possible: the mechanisms underlying the attribution of intentionality, agency and trust. As regards the theoretical and design strategy, we draw the model of such mechanisms from the biological theory of honest signalling. As Chella, Lebiere, Noelle, and Samsonovich (2011) hold it is likely that the conditions under which human and robotic agents successfully interact and pursue common goals are biologically inspired. Such conditions meet those that enable

* Corresponding author.
E-mail address: rosario.sorbello@unipa.it (R. Sorbello).

https://doi.org/10.1016/j.bica.2017.12.001
Received 3 August 2017; Received in revised form 28 November 2017; Accepted 1 December 2017
2212-683X/

Please cite this article as: Sorbello, R., Biologically Inspired Cognitive Architectures (2017), https://doi.org/10.1016/j.bica.2017.12.001

humans and animals to sense what is salient and act accordingly in a shared environment. Besides the theory of honest signalling has been already extended to human interaction laying the basis of sociometrics (Pentland, 2007). The paper is organized as follows. In the first section we reconstruct the biological meaning of the concept of honest signals. We emphasize the advantage of signalling intended as an automatic and perceivable communication that induces animals to choose stable strategies in competitive contexts. In the second section we present the extension of honest signalling to social human interactions and the project of social metrics. In the third section we describe the architecture that embeds the insight of sociometrics based on honest signalling and allows bringing in and controlling further conditions for a conceptual refinement.

2. Bio-inspired honest signalling theory

2.1. Honest signalling in biology: insights and stable state model

Zahavi (1975, 1977) and Zahavi and Zahavi (1999) brought the concept of honest signals in theoretical biology to account for cases in which individuals that compete with one another or have conflicting interests opt for a strategy that benefit them all rather than deceiving one another. Suppose that members of group A, which for instance belong to a prey species or are nestlings begging for food, and of group B, which belong to a predator species or are the feeding mother birds, have different access to information. Instead of following the incentive to cheat, A and B members shift to sharing information as the strategy that benefit both groups. To warrant the reliability of that communication they issue signals that cannot but being taken as honest because they are cost-added signals. Consider that gazelles and cheetahs are A and B members respectively. In the presence of cheetahs, gazelles make high up and down jumps instead of fleeing as if they wanted the predator to spot them. Thus gazelles show cheetahs to be able to flee by investing in a display of fitness, which is costly in terms of energy and time badly needed to run away. On the other hand, cheetahs learn that they cannot take the preys by surprise and may choose not to waste energy and time to hunt instead other preys. Furthermore, the gazelle that is able to jump that way shows cheetahs to have such strength that the predators will have to spend much more resources to try to catch her than those needed to chase another gazelle that is not able to display the same signal. Accordingly cheetahs use the added cost of this display as an observable gauge of signal reliability, because the cost of that jumping is much greater than the gain gazelles would get were it a phony signal. Therefore cheetahs can take the signal as honest communication that reduces uncertainty, because they come to know a quality that is possessed by some gazelles rather than others. In the case of nestlings and feeding mother birds, each nestling has an incentive to cheat and show it is hungrier than the other ones to receive food, while the mother bird has interest in knowing how much each nestling is hungry to feed the hungriest. Honest signals solve the problem of parent-offspring conflicting interest. Loud and harsh cries display hunger so that the louder and the harsher they are, the hungrier is the nestling that emits them. Those squawks have the further added cost that they may call the attention of predators. Therefore the risk of being caught outweighs the gain of cheating. If starving the nestling will bet against this risk and mother birds will get honest information about whether and how much each nestling is hungry. Signals of this kind are honest in a statistical sense. On average they show the receivers correctly the existence of an otherwise unobservable quality. They bear a cost that is added to that which the signallers undertake just to make sure that the signal is emitted with the physical properties needed to convey the information unambiguously. This is instead a strategic cost that means a reduction of fitness under some respects by which cheating or deceiving are constrained (Smith & Harper, 1995). In the case of prey-predator interaction, this signalling allows high fitness preys being distinguished from the other ones and deterring predation, thus serving predators to discriminate two subset in the preys group. Low fitness preys can't pretend to be otherwise because of the added cost that makes the signal unattainable for them. Honest signalling affects the behaviour of individuals by sharing and modifying the information to which they have access so that it increases their fitness. It leads to a state that has been qualified as stable by Grafen (1990) and Smith (1991).

We can summarize this result by letting:

1. *A* and *B* be any members of two competing or conflicting groups such that *A* has a two states quality, for instance hungry/satiated or strong/feeble, and *B* has a resource, be it food or deterred predation;
2. *p* be the probability that *A* is needy or strong and $(1-p)$ the inverse probability (for *B* these values have a uniform distribution);
3. *r* be a "coefficient of relatedness" such that a maximizing process of survival chance may pay B if it delivers the resource to *A* but if *A* is *p* (this measures the inclusive fitness for *A* and *B* being genetically related but it can be generalized to any case in which As and Bs pay offs are mutually dependent);
4. $(1-t)$ be the reduction factor in As survival chance due to the cost t of signalling that p.
5. $(1-d)$ be the reduction factor in Bs survival chance due to the cost d of delivering. *A* survival chance depends on his state and Bs delivering: if *A* is *p* and *B* delivers the chance is 1, if *A* is *p* and *B* does not deliver the chance is 0, if *A* is $(1-p)$ and *B* delivers the chance is 1, if *A* is $(1-p)$ and *B* does not deliver the chance is *X* with $0 < X < 1$. *B* survival chance is 1 if it does not deliver and $(1-c)$ otherwise. Given *A* and B are somehow related, the choice of each one benefits the other one r times. It can be shown that the equilibrium strategy occurs if the loss that reduces As survival chance, because *B* does not deliver while *A* is p, is greater or equal to the signalling cost t and the cost d, which reduces Bs survival chance, once weighted for r is greater or equal to the loss for *A* because *B* does not deliver while *A* is $(1-p)$. In such a case *A* will signal if p and *B* will deliver according to signalling. Instead if $(1-t)$ and $(1-d)$ weighted for r are not greater than the relatedness of *A* and *B*,*A* will never signal; if X and the relatedness of *A* and *B* are not greater than $(1-t)$ and $(1-d)$ weighted for r, *A* will always signal. On the other hand, if *B*s survival and X weighted for r are not greater than $(1-d)$ and the relatedness of *A* and *B*,*B* will always deliver; if $(1-d)$ and $(1-t)$ weighted for r is greater than B survival, *B* will never deliver. Therefore given an appropriate cost $t > 0$ honest signalling is a stable strategy in the sense that neither *A* nor *B* benefits from switching to another behaviour. Cheating and deceiving are still possible, rather they actually occur but this strategy is not optimal because it undermines the communication system that reduces uncertainty by sharing information and promotes the inclusive fitness of all parties.

2.2. From biology to sociometrics

Pentland (2008) has extended the theory of honest signalling to strategic contexts in which humans are engaged in face-to-face or group interactions. Speed dating and salary negotiations are examples of face-to-face interactions, while tactical decision making and coalition membership shifting within and across groups are examples of social aggregates interactions when conflicting or competing interests hold. Like in biology honest signals are unconscious, in the sense that they do not involve conscious reasoning, normative or linguistic judgments, mandatory and costly in terms of cognitive resources. Pentland (2008) describes four types of honest signals by which agents tune, synchronize or change cognitive features that are socially salient like attention, understanding, interest, focus and openness. The first type collects signals of the influence that agents have in the interaction, which is displayed by the distribution of attention to control and orienteer the communication and the behaviour. The second type collects mimicry signals that display the tuning of agents to each other, like nodding or