Author manuscript, published in "epirob (2009) Pages 201-202"

Towards a new social referencing paradigm

S. Boucenna[1], P. Gaussier[1,2], L. Hafemeister[1], K. Bard[3]
[1]ETIS, CNRS UMR 8051, ENSEA, Univ Cergy-Pontoise, [2]IUF,[3] Portmouth University
{boucenna,gaussier,hafemeister}@ensea.fr, kim.bard@port.ac.uk

How can a robot learn more and more complex tasks? This question is becoming central in robotics. In this work, we are interesting in understanding how emotional interactions with a social partner can bootstrap increasingly complex behaviors, which is important both for robotics application and understanding development. In particular, we propose that social referencing, gathering information through emotional interaction, fulfills this goal. Social referencing, a developmental process incorporating the ability to recognize, understand, respond to and alter behavior in response to the emotional expressions of a social partner, allows an infant, or a robot, to seek information from another individual and use that information to guide his behavior toward an object or event(Klinnert et al., 1983).

Gathering information through emotional interaction seems to be a fast and efficient way to trigger learning. This is especially evident in early stages of human cognitive development, but also evident in other primates (Russell et al., 1997). Social referencing ability might provide the infant, or a robot, with valuable information concerning the environment and the outcome of its behavior, and is particularly useful since there is no need for verbal interactions. In social referencing, a good object or event is identified or signaled with an emotional message, not with a verbal label. The emotional values can be provided by a variety of modalities of emotional expressions, such as facial expressions, voice, gestures, etc. We choose to use the facial expressions since they are an excellent way to communicate important information in ambiguous situations but also because they can be learned autonomously very quickly (Boucenna et al., 2008). Our idea is that social referencing as well as facial expression recognition can emerge from a simple sensori-motor system. All the work is based on the idea of the perception ambiguity: the unability at first to differentiate its own body from the body of others if they are correlated with its own actions. This perception ambiguity associated to a homeostatic system are sufficient to trigger first facial expression recognition and next learn to associate an emotional value to an arbitrary object. Without knowing that the other is an agent, the robot is able to learn some complex tasks. Hence we advocate the idea that the social referencing can be boostrapped from a simple sensori-motor system not dedicated to social interactions.

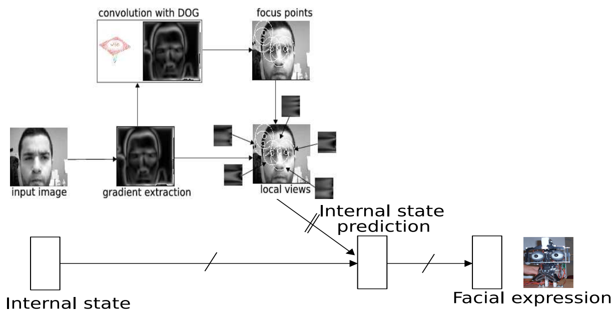

Figure 1: The global architecture to recognize facial expressions and imitate. A visual processing allows to extract the local views sequentially. The *internal state prediction* learns the association between the local views and the robot's internal state.

In our social referencing experiment, we have the following set-up: a robotic head is able to recognize the facial expressions, a Katana arm is able to reach an object in the workspace and an other camera views the workspace. Thanks to this set-up, the robot (head plus arm) can interact with the environment (human partner) and can manipulate objects. In the developed architecture, the robot learns to handle positive objects, and learns to avoid the negative objects as a direct consequence of emotional interactions with the social partner.

The robotic head learns to recognize emotional facial expressions (sadness, joy, anger, suprise and neutral face) autonomously (Boucenna et al., 2008). The internal emotional state of the robot triggers one specific expression and the human mimicks the robot face to face. The robot then learns to associate its internal emotional state with the human's facial expression. After few minutes of real time learning (typically less than 3 minutes) , the robot is able to recognize the human facial expressions as well as to mimick them (fig. 1).

After a visuo-motor learning, several positions in the workspace can be reached by the robot arm (Andry et al., 2001). One visual position corresponds to one or several motor configurations (e.g attractors). These attractors pull the arm in an attraction basin (the position target). This control

hal-00522716, version 1 - 1 Oct 2010

hal-00522716, version 1 - 1 Oct 2010

is performed with a dynamical system in the aim of smoothing the trajectory (Fukuyori et al., 2008). This dynamical system also uses a reinforcing signal in the aim of attaching a lot of or little importance to some attractors, for instance a reward can be given if the arm follows the right direction, otherwise a punition. The reinforcing signal can be emotional (e.g joy facial expression is a positive signal and an angry facial expression is a negative signal).

As soon as the facial expression learning is performed (i.e the human partner must imitate the robotic head between 2 and 3 min, then the robot is able to recognize and display the human facial expressions), the human can interact with the robotic head to associate an emotional value to an object (positive or negative). The neural network (N.N) processes in the same way signals from the robot's internal state and infomations correlated with this internal state (e.g facial expressions). The N.N does no distinction between the internal state and the facial expression recognized on the partner's face. In the absence of the internal state, the facial expression recognized induces an internal state which is associated with the object (a simple conditionning chain: figure 2).

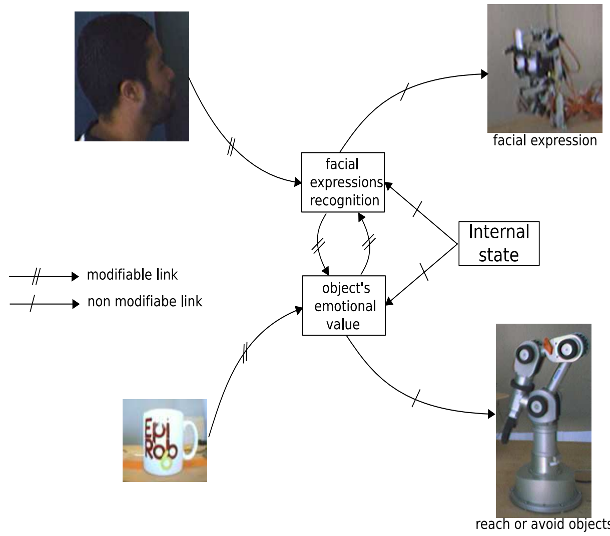

Figure 2: experimental set-up for the social referencing: The robot arm reachs the positive object and averts the negative object.

Classical conditioning can perform this association between the emotional value that the human transmits and some areas of the image. The attentional process used in this model is very simple, the robot focuses on colored patches or textures. When focusing on an object, the robot extracts some focus points and associates the recongition of the local view surrounding these focus points with the emotional value of the robot. For instance, if the robot is in a neutral emotional state and the human displays a joy facial expression in the presence of an object, the robot will move to a joy state and will associate a positive value to the object. On the contrary if the human displays an anger facial expression, the value associated to this object will be negative. As soon as this learning is finished, the robot arm can handle or avoid the objects according to their associated emotional value. In other words, the emotional value associated to the object is the reinforcing signal that the arm uses so as to move (exactly as the human facial expression or an internal reward). Besides, the human partner can always provide an emotional value on the robot's behavior, there is a competition between the object emotional value and the partner's facial expression. The competition is performed with a simple WTA (Winner Take All). The robot could handle negative objects if the human partner displays a joy facial expression or it could avert positive objects if the human displays an anger facial expression (the object's emotional value is not overwritten).

We think this approach can provide new interesting insights about how humans can develop social referencing capabilities from sensorimotors dynamics. In contrast to current developmental theory that social referencing is a complex cognitive process of triadic relations, the current work suggests 1) the primacy of emotion in learning, 2) the simple classical conditionning mechanisms by which anothers emotional signal assumes identity with internal emotional states, and 3) a simple system of pairing internal emotional state with object-directed behavior.

Acknowledgments

The authors thank: J. Nadel, M. Simon and R. Soussignan and P. Canet for the design and calibration of the robot head. L. Canamero for the interesting discussions on emotion modelling.This study is part of the European project "FEELIX Growing" IST-045169, the French Region Ile de France "DIGITEO"and the Institut Universitaire de France.

References

Andry, P., Gaussier, P., Moga, S., Banquet, J., and Nadel, J. (2001). Learning and communication in imitation: An autonomous robot perspective. *IEEE transactions on Systems, Man and Cybernetics, Part A*, 31(5):431–444.

Boucenna, S., Gaussier, P., and Andry, P. (2008). What should be taught first: the emotional expression or the face? *epirob*.

Fukuyori, I., Nakamura, Y., Matsumoto, Y., and Ishiguro, H. (2008). Flexible control mechanism for multi-dof robotic arm based on biological fluctuation. *From Animals to Animats 10*, 5040:22–31.

Klinnert, M., Campos, J., Sorce, J., Emde, R., and Svejda, M. (1983). The development of the social referencing in infancy. *Emotion in early development*, 2:57–86.

Russell, C., Bard, K., and Adamson, L. (1997). Social referencing by young chimpanzees (pan troglodytes). *journal of comparative psychology*, 111(2):185–193.