frontiers in
HUMAN NEUROSCIENCE

OPINION ARTICLE
published: 20 March 2012
doi: 10.3389/fnhum.2012.00057

Automatic human interaction understanding: lessons from a multidisciplinary approach

Anna Sedda[1]*†, Valentina Manfredi[2]†, Gabriella Bottini[1,3]†, Marco Cristani[4]† and Vittorio Murino[4]*†

[1] *Department of Humanistic Studies, University of Pavia, Pavia, Italy*
[2] *Fondazione IRCCS Istituto Neurologico "C. Besta," Milano, Italy*
[3] *Cognitive Neuropsychology Laboratory, Niguarda Ca' Granda Hospital, Milano, Italy*
[4] *Pattern Analysis and Computer Vision (PAVIS) Department, Istituto Italiano di Tecnologia, Genova, Italy*
**Correspondence: anna.sedda@unipv.it; vittorio.murino@iit.it*
†*Anna Sedda, Valentina Manfredi, Gabriella Bottini, Marco Cristani, and Vittorio Murino have contributed equally to this work.*

Humans are essentially a social species, as demonstrated by the fact that in everyday life people continuously interact with each other to achieve goals or simply to exchange states of mind (Frith, 2007; Frith and Frith, 2007; Adolphs, 2009). How people react to and interact with the surrounding world is a product of evolution: the success of our species is also due to our social intellect, allowing us to live in groups and share skills and purposes (Frith, 2007). In other words, our brain has evolved not only in terms of cognitive but also of social processing.

The "social brain" (Brothers, 1990) has the main goal of understanding and predicting what others are going to do next or, in other words, to figure out and predict others' intentions, which is an important task to interact successfully with the environment (Frith, 2007).

On one side, from its first introduction, the social brain has attracted much attention and in recent years neuroscientists have strongly focused on revealing mechanisms and brain areas involved in social processes (Adolphs et al., 1998; Damasio, 1998; Hari, 2003; Blakemore and Frith, 2004; Amodio and Frith, 2006; Frith, 2007; Frith and Frith, 2007; Adolphs, 2009; Hari and Kujala, 2009). Even though results are still preliminary, when it comes to understanding a social stimulus, four main actors have been identified to date: the amygdala, the temporal pole, the superior temporal sulcus, and the frontal cortices, particularly the medial prefrontal cortex, in its anterior and posterior rostral part and in the orbitofrontal area (Allison et al., 2000; Frith and Frith, 2006; Frith, 2007; Hari and Kujala, 2009).

On the other hand, social interactions are nowadays accessible to automatic analysis through computer science methods, namely, computer vision and pattern recognition (CVPR), the main disciplines used for automatic scene understanding (Turaga et al., 2008). In particular, social signal processing (SSP; Pentland, 2007; Vinciarelli et al., 2009) is a new research and technological area that aims at providing computers with the ability to sense and understand human *social* signals, i.e., signals produced during social interactions. Such signals are manifested through sequences of non-verbal behaviors including body posture, gesture, gaze and face expressions, and mutual distance (Vinciarelli et al., 2009). In addition, the pioneering advancements in SSP have shown that social signals, described as so elusive and subtle that only trained psychologists can recognize them, are actually evident and detectable enough to be captured by sensors like cameras, and interpreted through analysis techniques, typically derived by machine learning and statistics domains (Duda et al., 2000). Observation activities of social signals have never been as ubiquitous as today and they keep increasing in terms of both amount and scope. Furthermore, the involved technologies progress so much that some sensors already exceed human capabilities and, being easily available at a low cost, have an increasingly large diffusion.

However, the neuroanatomical correlates of social interaction have not been systematically shared with the SSP area due to the rare intersection of these disciplines. We aim to briefly review the most relevant methods for the automatic understanding of the social human behaviors from both the computational and the neuroscientific perspective, showing how they might gain large benefits from mutual interaction.

Behavioral indicators relevant for SSP come from researches in the emotional on the motor systems. Emotions in fact modulate and drive social interactions not only through facial expressions and prosodic vocalizations, that are traditionally investigated so far (Ekman, 1993; Adolphs et al., 1996; Anderson and Phelps, 1998; Fusar-Poli et al., 2009; Bonora et al., 2011), but also by means of *body language* (de Gelder et al., 2011). Interestingly, non-verbal behavior has mainly been studied by social sciences without a particular interest for the neurophysiological aspects of human interplays (Wolpert et al., 2003). The motor system plays indeed a pivotal role in social cognition, as motor predictive mechanisms may contribute to the anticipation of what others are going to do next and regulate our own reactions, a principal function of social cognition (Wolpert et al., 2003; Frith and Frith, 2007; Adolphs, 2009; Hari and Kujala, 2009). Revealingly, the mirror system, which has been shown first to operate for motor acts (Rizzolatti and Craighero, 2004), has now been dragged into the discussion also for the processing of social stimuli (Frith and Frith, 2007). The mirror system is regarded as the basis for shared motor representations between the producer and the recipient of a motor act-based message (Rizzolatti and Craighero, 2004). Analogously, it has been suggested that when we need to read a hidden intention or emotional state of others during an interaction we activate a similar pattern in our brain areas, sharing the feeling of the interlocutor to understand it (Wicker et al., 2003; Wolpert et al., 2003; Frith, 2007).

Some authors do not believe that perception of complex states of mind could be inferred only by observing an action (Jacob and Jeannerod, 2005). It is true that the same action, e.g., grasping a knife, could lead to two different scenarios: an aggression or the cutting of an apple (Jacob and Jeannerod, 2005). Nevertheless the environment in which an action occurs may significantly influence the comprehension of the

intention of the action itself. In the case of automatic processing of human behavior, the detection of a person grasping a knife in an environment such as an airport would be in any case a signal of danger. Although the real intentions cannot be read using only motor gestures (de Gelder et al., 2011), it is clear that for some practical applications it is sufficient to detect specific occurring events, but it would be even more important to *prevent* a dangerous situation even at the cost of some false alarms. Furthermore, recent evidences suggest that Jacob and Jeannerod critique may not be correct, as several studies demonstrate that, even in absence of context information, intentions translate into differential kinematic patterns (Becchio et al., 2008a,b; Sartori et al., 2009) and observers are especially attuned to kinematic information, and might use early differences in visual kinematics to anticipate the intention of an agent in performing a given action (Manera et al., 2011; Sartori et al., 2011).

The common ground of SSP and studies of emotions should be to adapt the automatic systems for monitoring and surveillance to cerebral systems human interactions. More specifically, the ongoing trend of approaching monitoring scenarios with SSP methods is strongly motivated by the fact that social signals are now starting to be considered as stable, reliable, and genuine traits of the behavioral state of a person (de Gelder et al., 2011). Similarly, this same logic guided recent advances in the interaction between humans and machines (Tao and Tieniu, 2005). In other words, human behavior is now considered as a phenomenon subjected to rigorous principles that produces predictable patterns of activities, and that humans use social signals to convey, *often outside conscious awareness*, their attitude toward other people and social environments, as well as emotions (Richmond and McCroskey, 1995).

Consequently, understanding the processes underlying human behavior in social interactions starting from motor gestures and other social cues is extremely important to design automatic systems able to model specific situations and events in a principled way. This can be faced by capturing novel features (e.g., specific postures, subtle gestures, mutual distances) which have a precise meaning as consequences of activations of well defined parts of the brain network (comprising the prefrontal parietal and temporal areas; Wolpert et al., 2003). Moreover, motor gestures could be the only objective indicators of emotional behavior, although they do not allow mind reading (e.g., knowing in advance that a person will hit somebody because he has psychiatric problems rather than because he has been offended), rather to anticipate that a social action will take place (e.g., somebody will be hit).

The systematic investigation of basic emotional gestures has provided databases of bodily expressive postures (Atkinson et al., 2004; de Gelder and Van den Stock, 2011; de Gelder et al., 2011). These databases have been developed using actors displaying emotions categorized through forced choice paradigms (Winters, 2005).

More information about the neural systems involved in predicting and decoding human interactions might be derived from monitoring cerebral activity while subjects watch video sequences of people interacting in ecological contexts. The main difference between this approach and traditional studies would be using complex interactions in the ecological context rather than single postures as stimuli. In this way, computational algorithms would benefit from indicators validated by neurological pattern activations, that are discovered using ecological interactions, thus allowing one to recognize with a greater accuracy bodily expressions in complex real scenarios. Consequently, the classical CVPR approach of learning by examples can be safely utilized due to the support by a reliable neuroscientific basis. Furthermore, using non-invasive brain techniques, such as transcranial magnetic stimulation, it could be possible to confirm the brain areas involved in social interaction processing, clarifying dissociations, and whether these circuits are really needed or only implicated in this process, as it has occurred in other neuroscience domains (Ellison et al., 2004).

The use of fMRI or TMS would also allow to detail the involvement of different cerebral regions in different body expressions (de Gelder and Van den Stock, 2011). Moreover it could also be predicted that the initial hand and arm position and velocity could indicate an aggression. Studying emotional value of body expressions could benefit from more advanced technologies also able to record movements velocity (Wolpert et al., 2003) not only assuming the (possibly) wrong perspective of imitations (Jacob and Jeannerod, 2005). This theoretical approach would be similar to that used to categorize facial expressions (Darwin, 1872; Ekman and Friesen, 1969). Moreover, spontaneous dynamic expressions could help in confirming the neural basis of emotional body postures, so far only obtained through elicited stimuli (de Gelder et al., 2011).

In this way, neuroscience knowledge, resulting from neuroimaging and behavioral experiments, could provide SSP with reliable indicators of human behaviors being helpful to identify and predict events of interest. A deeper understanding of the neural circuits underpinning social interactions could be useful for SSP because it would provide a stronger evidence that the behavioral indicators taken into account by automatic analyses systems are the correct ones, or in other words are those that also the "real" brain uses. Computer science, in turn, could provide automatic computational techniques useful to better analyze single or sequences of action units. In particular, methods for gesture decoding, for the scrutiny of body postures, and for the extraction of proxemic cues are only a few examples of the technology. In this way, the video modality could be finally considered extensively in the analysis, whereas the audio channel has been traditionally the most used information source by neuroscientists so far.

In conclusion, to empower the available methodologies, more intersection between Neuroscience and SSP is needed to construct a more unitary frame of research for a better understanding of human behaviors through the study of the emotional and the motor system. Indeed, understanding the processes underlying human behavior in social interactions is extremely important to design systems able to detect, recognize, or, better, model, and predict specific situations and events in an automatic fashion.

REFERENCES

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* 60, 693–716.

Adolphs, R., Damasio, H., Tranel, D., and Damasio, A. R. (1996). Cortical systems for the recognition of emotion in facial expressions. *J. Neurosci.* 16, 7678–7687.

Adolphs, R., Tranel, D., and Damasio, A. R. (1998). The human amygdala in social judgment. *Nature* 393, 470–474.

Allison, T., Puce, A., and Mccarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci. (Regul. Ed.)* 4, 267–278.

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277.

Anderson, A. K., and Phelps, E. A. (1998). Intact recognition of vocal expressions of fear following bilateral lesions of the human amygdala. *Neuroreport* 9, 3607–3613.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., and Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33, 717–746.

Becchio, C., Sartori, L., Bulgheroni, M., and Castiello, U. (2008a). Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition* 106, 894–912.

Becchio, C., Sartori, L., Bulgheroni, M., and Castiello, U. (2008b). The case of Dr. Jekyll and Mr. Hyde: a kinematic study on social intention. *Conscious. Cogn.* 17, 557–564.

Blakemore, S. J., and Frith, U. (2004). How does the brain deal with the social world? *Neuroreport* 15, 119–128.

Bonora, A., Benuzzi, F., Monti, G., Mirandola, L., Pugnaghi, M., Nichelli, P., and Meletti, S. (2011). Recognition of emotions from faces and voices in medial temporal lobe epilepsy. *Epilepsy Behav.* 20, 648–654.

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* 1, 27–51.

Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous system. *Brain Res. Brain Res. Rev.* 26, 83–86.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals.* London: John Murray.

de Gelder, B., and Van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Front. Psychol.* 2:181. doi: 10.3389/fpsyg.2011.00181

de Gelder, B., Van den Stock, J., Meeren, H. K., Sinke, C. B., Kret, M. E., and Tamietto, M. (2011). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neurosci. Biobehav. Rev.* 34, 513–527.

Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification,* 2nd Edn. New York: Wiley.

Ekman, P. (1993). Facial expression and emotion. *Am. Psychol.* 48, 384–392.

Ekman, P., and Friesen, W. V. (1969). A tool for the analysis of motion picture film or video tape. *Am. Psychol.* 24, 240–243.

Ellison, A., Schindler, I., Pattison, L. L., and Milner, A. D. (2004). An exploration of the role of the superior temporal gyrus in visual search and spatial perception using TMS. *Brain* 127, 2307–2315.

Frith, C. D. (2007). The social brain? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 671–678.

Frith, C. D., and Frith, U. (2006). How we predict what other people are going to do. *Brain Res.* 1079, 36–46.

Frith, C. D., and Frith, U. (2007). Social cognition in humans. *Curr. Biol.* 17, R724–R732.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., Barale, F., Perez, J., Mcguire, P., and Politi, P. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J. Psychiatry Neurosci.* 34, 418–432.

Hari, R. (2003). Neural basis of social cognition. *Duodecim* 119, 1465–1470.

Hari, R., and Kujala, M. V. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiol. Rev.* 89, 453–479.

Jacob, P., and Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends Cogn. Sci. (Regul. Ed.)* 9, 21–15.

Manera, V., Becchio, C., Cavallo, A., Sartori, L., and Castiello, U. (2011). Cooperation or competition? Discriminating between social intentions by observing prehensile movements. *Exp. Brain Res.* 211, 547–556.

Pentland, A. (2007). Social signal processing. *IEEE Signal Process. Mag.* 24, 108–111.

Richmond, V., and McCroskey, J. (1995). *Nonverbal Behaviors in Interpersonal Relations.* Boston: Allyn and Bacon.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Sartori, L., Becchio, C., Bara, B. G., and Castiello, U. (2009). Does the intention to communicate affect action kinematics? *Conscious. Cogn.* 18, 766–772.

Sartori, L., Becchio, C., and Castiello, U. (2011). Cues to intention: the role of movement information. *Cognition* 119, 242–252.

Tao, J., and Tieniu, T. (2005). Affective computing: a review. Affective computing and intelligent interaction. *Lect. Notes Comput. Sci.* 3784, 981–995.

Turaga, P., Chellappa, R., Subrahmanian, V., and Udrea, O. (2008). Machine recognition of human activities: a survey. *IEEE Trans. Circuits Syst. Video Technol.* 18, 1473–1488.

Vinciarelli, A., Pantic, M., and Bourlard, H. (2009). Social signal processing: survey of an emerging domain. *Image Vis. Comput.* 27, 1743–1759.

Wicker, B., Perrett, D. I., Baron-Cohen, S., and Decety, J. (2003). Being the target of another's emotion: a PET study. *Neuropsychologia* 41, 139–146.

Winters, A. (2005). Perceptions of body posture and emotion: a question of methodology. *N. Sch. Psychol. Bull.* 3, 35–45.

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 593–602.

Received: 03 February 2012; accepted: 02 March 2012; published online: 20 March 2012.

Citation: Sedda A, Manfredi V, Bottini G, Cristani M and Murino V (2012) Automatic human interaction understanding: lessons from a multidisciplinary approach. Front. Hum. Neurosci. ***6****:57. doi: 10.3389/fnhum.2012.00057*