

You and I, robot

Shaun Gallagher

Received: 7 July 2011 / Accepted: 30 March 2012 / Published online: 13 April 2012
© Springer-Verlag London Limited 2012

Abstract I address a number of issues related to building an autonomous social robot. I review different approaches to social cognition and ask how these different approaches may inform the design of social robots. I argue that regardless of which theoretical approach to social cognition one favors, instantiating that approach in a workable robot will involve designing that robot on enactive principles.

Keywords Social robots · Social cognition · Theory of mind · Theory theory · Simulation theory · Interaction theory · Enactive cognition

1 Introduction

Researchers in advanced robotics are attempting to build autonomous social robots that will be able to seamlessly and reliably interact with humans in specific situations.¹ This project motivates both philosophical and practical questions about what precisely is required in a robot if it is to be able to engage in something close to the kind of interaction that characterizes human–human relations, even if only on a pragmatic level. That is, if we set aside concerns that have to do with the complicating factors of care and emotion and focus simply on communicative capacity, is there some guiding ideal (in the sense of a Kantian ideal

that we could aim at, even if we are not convinced that we can achieve it) for designing and building such a robot?

In setting aside questions about care and emotion I don't mean to suggest that these issues are not important for human or human–robot interactions or that they are not solvable. Even now it seems possible to build robots that elicit care and certain emotions from humans (e.g., Kismet, see Breazeal 2002). Even if the robot is not designed, in terms of its appearance and behavior (e.g., facial features, vocal intonations, etc.), to elicit care from a human responder, it seems possible that humans may come to care for a robot in a significant way to the extent that they come to depend on it. This sometimes happens in regard to other machines—automobiles, computers, etc. It's an open question, however, whether this kind of caring for a machine is the same kind of caring, and simply a matter of a difference of degree from that which humans experience for each other. It's also an open question whether caring can go the other way, that is, whether a robot can have anything more than a pragmatic care for a human (in the sense of simply taking care of that human). Also, setting these questions aside doesn't mean that care and emotion do not enter into and shape our everyday human interactions in important ways. Stripping away care and emotion from our everyday interactions may in fact change them in

S. Gallagher (✉)
Lillian and Morrie Moss Chair of Excellence in Philosophy,
University of Memphis, Memphis, TN, USA
e-mail: s.gallagher@memphis.edu

S. Gallagher
Cognitive Science, School of Humanities,
University of Hertfordshire, Hatfield, UK

¹ I'm involved in a large project of this sort, although my contribution is in the distant theoretical background and is focused on questions about the nature of gesture and the possibility of building gesture into the repertoire of a robot's communicative skills. My research on robotics is supported by a grant from the Robotics Collaborative Technology Alliance and General Dynamics, #64018180, Social cues and behaviors in HR collaboration. Also, thanks to the Marie Curie Initial Training Network, Towards an Embodied Science of Intersubjectivity (TESIS). Marie Curie Actions, European Commission Research for support of my research on intersubjectivity.