

## Trusting Humans and Avatars: An fMRI Study

René Riedl, Peter N. C. Mohr, Peter H. Kenning, Fred D. Davis, Hauke R. Heekeren

Corresponding author: rene.riedl@jku.at

Information and communication technologies have dramatically changed the context in which interactions among humans take place. In particular, social interactions have become increasingly more often computer-mediated. As a result of this shift from face-to-face conversation to Internet-based communication, traditional human-human interactions are becoming increasingly more often computer-mediated. Using computers and the Internet, however, usually increases a perception of uncertainty. A major reason for this fact is that an interaction partner's face, which is typically not observable in computer-mediated interaction, serves the interpersonal function of allowing one person to predict another's personality traits and behavior (Oosterhof and Todorov 2008), which has been demonstrated to apply, particularly, to trustworthiness predictions (Todorov 2008; Winston et al. 2002). Thus, prediction of another person's personality traits and behavior is more difficult in computer-mediated interaction than in traditional face-to-face interaction.

To overcome the perception of uncertainty in online environments, interactions on the Internet occur increasingly more often via virtual humans who possess realistic faces, so-called avatars (Donath 2007). Because attention to a realistic portrayal of a human face is of particular importance for trust perceptions (Todorov et al. 2008), it has been theorized that the use of avatars may positively affect trust perceptions in computer-mediated interactions, thereby mitigating uncertainty perceptions. In fact, recent research found that avatar-based communication increases perceived social presence and interpersonal trust, although an avatar's face does not need to resemble a computer user's real face because it can be chosen freely (Bente et al. 2008).

Against the background of the increased use of avatars as instruments for trustful online interaction, it is rewarding to study both the similarities and differences in trust behavior and underlying neural mechanisms between human-human and human-avatar interactions. To date, neurobiological research has already revealed insights into the genetical, hormonal, and brain mechanisms of trust (for a review on the biology of trust, see Riedl and Javor 2012). Most of these studies have used two-person exchange games, in particular the trust game, to reveal those insights. In this game trust is based on an individual's inferences about the intentions of the other player to predict his or her trustworthiness (Berg et al. 1995).

The ability to infer the internal states of other actors to predict their traits, intentions, and behavior is known as theory of mind, and the underlying inference process is commonly referred to as mentalizing (e.g., Amodio and Frith 2006). Hence, theory of mind and mentalizing are fundamental concepts in trust situations because the decision to trust involves thinking about an interaction partner's traits, intentions, and behavior. In fact, neuroimaging studies based on the trust game found significant activation in the medial frontal cortex (for a review, see Riedl and Javor 2012).

We studied individuals' trust behavior and underlying brain mechanisms in a multi-round trust game setting. Participants played in the role of the investor against both humans and avatars. The findings reveal that a brain structure associated with theory of mind (mentalizing), the medial frontal cortex, was significantly more activated in the human versus avatar condition. Moreover, we found this enhanced mentalizing activity to be related to elevated abilities to discriminate between trustworthy and untrustworthy actors. However, we did not find a difference in participants' trustworthiness learning rates.

### REFERENCES

- ❖ Amodio, D. M., and Frith, C.D. 2006. "Meeting of minds: The medial frontal cortex and social cognition," *Nature Reviews Neuroscience* (7), pp. 268-277.
- ❖ Bente, G., Rüggenberg, S., Krämer, N. C., and Eschenburg, F. 2008. "Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations," *Human Communication Research* (34), pp. 287-318.
- ❖ Berg, J., Dickhaut, J., and McCabe, K. 1995. "Trust, reciprocity, and social history," *Games and Economic Behavior* (10), pp. 122-142.
- ❖ Donath, J. 2007. "Virtually trustworthy," *Science* (317), pp. 53-54.
- ❖ Oosterhof, N. N., and Todorov, A. 2008. "The functional basis of face evaluation," *Proceedings of the National Academy of Sciences* (105), pp. 11087-11092.
- ❖ Riedl, R., and Javor, A. 2012. "The biology of trust: Integrating evidence from genetics, endocrinology and functional brain imaging," *Journal of Neuroscience, Psychology, and Economics* (5), pp. 63-91.
- ❖ Todorov, A. 2008. "Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach / avoidance behaviors," *Annals of the New York Academy of Sciences* (1124), pp. 208-224.
- ❖ Todorov, A., Said, C. P., Engell, A. D., and Oosterhof, N. N. 2008. "Understanding evaluation of faces on social dimensions," *Trends in Cognitive Sciences* (12), pp. 455-460.
- ❖ Winston, J. S., Strange, B. A., O'Doherty, J., and Dolan, R. J. 2002. "Automatic and intentional brain responses during evaluation of trustworthiness of faces," *Nature Neuroscience* (5), pp. 277-283.