The Uncanny Valley: Making Human-Nonhuman Distinctions

Julie Carpenter (julie4@u.washington.edu)
University of Washington, Department of Technical Communication, 14 Loew Hall, Box 352195
Seattle, WA 98195 USA

Matthew Eliot (mjeliot@u.washington.edu)
University of Washington, Department of Technical Communication, 14 Loew Hall, Box 352195
Seattle, WA 98195 USA

Daniel Schultheis (dschultheis@gmail.com)
University of Washington, Department of Mathematics, Box 354350 and Department of Physics, Box 351560
Seattle, WA 98195 USA

Abstract

Humans bring a complex infrastructure of beliefs and expectations to their social relationships which can drive both social action and social thought. As robots become increasingly anthropomorphic, the potential for humans to bring this same infrastructure to human-robot interaction increases as well. Mori's Uncanny Valley theory can serve as an interesting lens through which to explore one of humanity's most primal social drives: the ability to make clear human-nonhuman distinctions.

Introduction

Human's relationships with one another can be characterized as being based in the exchange of resources (Homans, 1961). We offer one another material goods, time, kinship, labor, attention, and information – this is the premise of Social Exchange Theory, as originally proposed by Gouldner (1960). Our methods of exchange, however, are emblematic of the deeper human drives that motivate resource exchange. Anthropologist Alan Fiske (1991) suggests that humans engage in four basic means for resource exchange:

- *Communal Sharing* reveals the human drive for group identity.
- *Authority Ranking* describes the need to determine one's place in the social hierarchy.
- *Equality Matching* connotes our need for reciprocity in our relationships.
- *Market Pricing* expresses the human tendency to assign relative value.

Fiske calls these methods for exchanging resources "relational models," for they are each founded on primal human social needs that drive the formation, maintenance, and termination of social relationships.

We suggest that each of these relational models can deeply affect the ways in which humans conceptualize and relate to androids, or social humanoid robots. In this poster, we will focus specifically on the important features of Communal Sharing as a lens through which to better understand the phenomenon known as The Uncanny Valley.

Reeves & Nass (1996) posited that humans treat computers as social actors, projecting human emotions and intentions onto these machines. The visual design of humanoid robots, which are essentially embodied computers, may intensify the human tendencies to apply the same "rules" for human-human interactions with these robots. Mori (1970) described a more problematic reaction, a type of cognitive dissonance (Draycott & Drabbs, 1998) in which conflicting sets of expectations can lead humans to experience strong feelings of discomfort and uncertainty when interacting with highly humanoid robots. Dubbing this phenomenon the "Uncanny Valley," Mori describes this dissonance as a clash between expectations related to appearance and familiarity:

> *"In this case (of a prosthetic human hand), there is no longer a sense of familiarity. It is uncanny. In mathematical terms, strangeness can be represented by negative familiarity, so the prosthetic hand is at the bottom of the valley. So in this case, the appearance is quite humanlike, but the familiarity is negative. This is the uncanny valley."*

Communal Sharing

Fiske's (1991) conception of Communal Sharing as a relational model is based on the human drive to determine social group membership, often at the expense of individual identity. Humans are driven to draw distinctions between those who belong and those who do not. Once these distinctions are made, the group implicitly or explicitly confers certain rights and privileges to group members, such as the following:

- Group members are often accorded equal opportunities within the group's social sphere.
- Group members can give and take freely from pooled resources.

- Decision-making is done by consensus.
- Group members seek uniformity and conformity with one another.
- Responsibilities are viewed collectively, with everyone expected to contribute as needed.

Communal Sharing, as a relational model, captures the human drive to form egalitarian groups. In such cases, the identification of fellow group members serves to indicate to others who are familiar with the protocols and practices of that particular group as well as to protect group resources from outsiders.

The Uncanny Valley as a Social Phenomena

Membership in the human race is itself a type of group membership, which has culturally determined benefits as well as disadvantages for those who do not clearly qualify (Goffman, 1963). Anthropomorphism – the tendency to attribute human motivations, characteristics, or other traits to non-human things – is somewhat easily triggered. For example, in Mori's "human hand" example quoted earlier, the humanlike appearance of a prosthetic hand creates dissonance because it is difficult for a person to align the artificial hand's appearance with unfamiliarity. Mueller (2004) states that there is a "persistent tendency for humans to anthropomorphize, that is, to attribute human qualities to inanimate objects"; a simple human like shape can be enough to generate a set of complex emotional responses. With anthropomorphic design rampant in social robots, what happens when humans cannot make clear distinctions between human and non-human android? In the Communal Sharing relational model, humans would be drawn to treat the android as if it were human: extending concern, assuming shared values, and expecting appropriate social behavior. This tendency would conflict with the perception that they are interacting with a machine, and the Uncanny Valley as a social phenomenon is born.

In such moments of cognitive dissonance, humans experience a divergence of expectations about the types of interaction and exchange available. Duffy (2003) suggests that human acceptance of social robots hinges on the successful leveraging of either set of such expectations:

> *"If one bootstraps these expectations, i.e., through exploiting anthropomorphism, one can be more successful in making social robots less frustrating to deal with and perceived as more helpful."*

The Communal Sharing relational model suggests that humans could be more comfortable in human-android interactions by developing a specific social category for androids that clearly delineates how to treat such humanlike creatures, which would include appropriate levels of expectation. As the level of anthropomorphism in androids increases and the urge to treat androids as members of the human race increases as well, humans will need to develop such social schemata for sustained interaction.

Conclusion

In this poster, we have proposed exploring one part of Social Exchange Theory, Communal Sharing, to further leverage android design towards more comfortable, natural human-robot interactions. While not all android roles or services will necessarily require mimicked human-human interaction protocol, we suggest that based on previous research discussed here, it is human tendency to (a) approach human-robot interaction with very individual expectations based on personal experience and (b) assign humanlike attributes to androids. When these expectations are not met or another factor creates cognitive dissonance, the success of the interaction is diminished. However, this will not be a static set of schemata; as androids become more pervasive in society, and are used for different roles and tasks, user expectations and preferences will adjust accordingly.

References

Draycott, S. & Drabbs, A. (1998). Cognitive dissonance: An overview of the literature and its integration into theory and practice of clinical psychology. *The British Journal of Clinical Psychology*; Sep 1998; 37 (3), 341-354.

Duffy, Brian. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42: 177-19.

Fiske, A. (1991). *Structures of social life*. New York: Free Press.

Goffman E. 1963. *Stigma: Notes on the management of spoiled identity*. Englewood Cliffs, NJ: Prentice Hall.

Gouldner, A.W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25, 161-178.

Homans, G. (1961). *Social behavior*. New York: Harcourt, Brace & World.

Mori, M. (2005). The Uncanny Valley. *Energy*, 7(4), 33-35. (K.F. MacDormand & T. Minato, Trans.). *Energy*, 7, 33-35. (Original work published 1970).

Mueller, J. (2004). Review essay: Emotional design by Donald A. Norman. *International electronic journal for leadership in learning.* (8), 14, Dec. 17. University of Calgary Press: Canada.

Reeves, B. & Nass, C. *The media equation*. CSLI Publications, Stanford, 1996.