Context Area:
Human Interaction with Autonomous Entities

Examiner: Bradley Rhodes
Research scientist, Ph.D.
Ricoh Innovations, Inc., Menlo Park, CA

Description

I am interested in how humans relate to the autonomy of intelligent artificial entities, such as software agents and autonomous robots. How would humans react to and interact with such intelligent non-human autonomous entities? From the human perspective, what level of autonomy is appropriate, expected, and useful? How much control would humans like to have over increasingly intelligent, e.g., context sensitive and adaptive, autonomous entities?

Areas that could help to answer such questions might be related to software agents, avatars, and autonomous robots. However, the focus of my question is not the technological and architectural details of these systems, but what humans want, don't want, expect, and so forth. Since there are not many examples of highly intelligent autonomous entities yet, it is not about the social consequences of current technologies, but the social consequences of the future introduction of these new technologies.

Limitations: This is my contextual area, so it is:
- Not about technology details
- Not about architecture of agents or robots
- Not about autonomy itself (or how to do that, technically), but the influence of it on people and society
- Not about interaction with dumb but autonomous technologies (e.g., air-conditioning)
- Although there are no such autonomous artificial entities yet, I assume that they will be created in the future: how humans will deal with them?

Written Requirement
The written requirement for this area will consist of a 24-hour take-home exam.

Signature: _____   Date: _____

Reading list

The reading list is structured in three sub areas:

- **Sociological and psychological aspects of interactions with autonomous systems**
  - Human expectations towards autonomous entities/systems/agents
  - Social responses (analog to Computers as Social Actors)
  - Society and autonomous entities
  - Autonomy and "Aliveness" of objects
- **User interface design issues**
  - Adjustable Autonomy
  - Interface design for autonomous systems
  - Human-centered autonomous systems
  - Advanced human-robot relations
  - Function allocation between agents (humans and machines) in a sociotechnical system
- **Case studies of social interactions between humans and autonomous entities**
  - Software agents, specifically socially intelligent agents (SIA)
  - Robots, specifically socially intelligent autonomous robot (SIAR)
  - Avatars

The online version of this reading list contains links to most of the papers, as well as abstracts for most items: **http://www.media.mit.edu/~stefanm/generals/**

Sociological and psychological aspects of interactions with autonomous systems

Donald A. Norman (1994). *How Might People Interact with Agents*. Communications of the ACM 37 (7), July 1994, pp. 68-71.

Jonathan Steuer (1995). *Self vs. Other; Agent vs. Character; Anthropomorphism vs. Ethopoeia.* In *Vividness and Source of Evaluation as Determinants of Social Responses Toward Mediated Representations of Agency*, doctoral dissertation, Stanford U, advised by Nass and Reeves.

Lars Oestreicher, Helge Hüttenrauch, and Kerstin Severinsson-Eklund (1999). *Where are you going little robot? – Prospects of Human-Robot Interaction.* Position paper for the CHI '99 Basic Research Symposium.

Valentino Braitenberg (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge MA: The MIT Press.

K. Bumby and Kerstin Dautenhahn (1999). *Investigating Children's Attitudes Towards Robots: A Case Study.* Proceedings of CT99, The Third International Cognitive Technology Conference, August, 1999, San Francisco CA.

Kerstin Dautenhahn (1998). *The Art of Designing Socially Intelligent Agents – Science, Fiction, and the Human in the Loop***.** Special Issue *Socially Intelligent Agents*, Applied Artificial Intelligence Journal, Vol. 12, 7-8, pp. 573-617.

Cynthia Breazeal (1999). *Robot in Society: Friend or Appliance?* In Agents99 Workshop on Emotion-Based Agent Architectures, Seattle, WA, pp. 18-26.

David Stork (ed.) (1997). *HAL's legacy: 2001's computer as dream and reality*. Cambridge MA: The MIT Press, chapters 1, 2, and 9.

Clifford Nass, Steuer, J., Tauber, E., and Reeder, H. (1993). *Anthropomorphism, Agency, & Ethopoeia: Computers as Social Actors*. Presented at INTERCHI '93; Conference of the ACM / SIGCHI and the IFIP; Amsterdam, Netherlands, April 1993.

Kerstin Dautenhahn (2000). Soc*ially Intelligent Agents and The Primate Social Brain - Towards a Science of Social Minds*. Proceedings of AAAI Fall Symposium *Socially Intelligent Agents - The Human in the Loop*, AAAI Press, Technical Report FS-00-04, pp. 35-51.

Kerstin Dautenhahn (1999). *Embodiment and Interaction in Socially Intelligent Life-Like Agents*. In C. L. Nehaniv (ed.) *Computation for Metaphors, Analogy and Agent*, Springer Lecture Notes in Artificial Intelligence, Volume 1562, New York, NY: Springer, pp. 102-142.

Robert D. Putnam (2000). *Bowling alone: The Collapse and Revival of American Community*. New York, NY: Simon and Schuster, selected chapters.

Robert D. Putnam (1995). *Bowling Alone: America's Declining Social Capital*. Journal of Democracy 6:1, January 1995, pp. 65-78.

Douglas R. Hofstadter and Daniel C. Dennett (1981). *The Mind's I: Fantasies and Reflections on Self and Soul*. New York, NY: Basic Books, chapters 4, 5, 8, 10, 11, 13, 18, 22.

Byron Reeves and Clifford Nass (1996). *The Media Equation*. Stanford, CA: Cambridge University Press, selected chapters.

Anne Foerst (1995). *The Courage to Doubt: How to Build Android Robots as a Theologian*. Talk, presented at Harvard Divinity School, November 27, 1995.

Joseph Weizenbaum (1976). *Computer power and human reason: From judgment to calculation*. San Francisco, CA: W.H. Freeman, pp. 1-16; 202-227; 258-280.

Joseph Weizenbaum (1966). *ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine*. Communications of the ACM 9(1):36-45.

Daniel C. Dennett (1987). *The Intentional Stance*. Cambridge, MA: The MIT Press.

Bill Joy (2000). *Why The Future Doesn't Need Us.* Wired Magazine 8.04.

Bruce Tognazzini (1994). *STARFIRE: A Vision of Future Computing* (video).

Erik Brynjolfsson and Michael Smith (2000). *The Great Equalizer? Customer Choice Behavior at Internet Shopbots*. Unpublished paper.

User interface design issues

Dennis Perzanowski, A. Schultz, E. Marsh, and W. Adams (2000). *Two Ingredients for My Dinner with R2D2: Integration and Adjustable Autonomy.* Papers from the 2000 AAAI Spring Symposium Series, Menlo Park, CA: AAAI Press.

Rino Falcone and Cristiano Castelfranchi (2000). *Levels of Delegation and Levels of Adoption as the basis for Adjustable Autonomy.* Lecture Notes in Artificial Intelligence n°1792, pp. 285-296.

Michael Mogensen (2001). *Dependent Autonomy and Transparent Automatons?* In Lars Qvortrup (ed.) *Virtual Interaction: Interaction in/with Virtual Inhabited 3D Worlds,* New York, NY: Springer.

Dennis Peraznowski, William Adams, Alan Schultz, and Elaine Marsh (2000). *Towards Seamless Integration in a Multi-modal Interface.* Workshop on Interactive Robotics and Entertainment, Carnegie Mellon University: AAAI Press, pp. 3-9.

Eric Horvitz (1999). *Principles of Mixed-Initiative User Interfaces.* ACM CHI'99 Proceedings, pp. 159-166.

Ben Shneiderman (1997). *Direct Manipulation for Comprehensible, Predictable, and Controllable User Interfaces.* Proceedings of IUI97, International Conference on Intelligent User Interfaces, Orlando, FL, January 6-9, pp. 33-39.

Marc Mersiol, Ayda Saidane (2000). *A Tool to Support Function Allocation.* Proceedings of Safety and Usability Concerns in Aeronautics, SUCA 2000.

Gregory A. Dorais, R. Peter Bonasso, David Kortenkamp, Barney Pell, and Debra Schreckenghost (1998). *Adjustable Autonomy for Human-Centered Autonomous Systems on Mars.* Proceedings of the First International Conference of the Mars Society, Aug. 1998.

Alan Wexelblat and Pattie Maes (1997). *Issues for Software Agent UI.* Unpublished paper.

Ben Shneiderman and Pattie Maes (1997). *Direct manipulation vs. interface agents. Excerpts from debates at IUI 97 and CHI 97.* interactions, 4(6):42-61.

Examples of social interactions between humans and autonomous entities

Dennis Perzanowski, A. Schultz, W. Adams, and E. Marsh (2000). *Using a Natural Language and Gesture Interface for Unmanned Vehicles.* In Unmanned Ground Vehicle Technology II,

G.R. Gerhart, R.W. Gunderson, C.M. Shoemaker (eds.), Proceedings of the Society of Photo-Optical Instrumentation Engineers, vol. 4024, pp. 341-347.

Phoebe Sengers, Simon Penny, and Jeffrey Smith (2000). *Traces: Semi-Autonomous Avatars.* Unpublished paper.

Kerstin Dautenhahn (1999*). Robots as Social Actors: AURORA and the Case of Autism*. Proceedings of CT99, The Third International Cognitive Technology Conference, August 1999, San Francisco, CA, pp. 359-374.

Milind Tambe, David V. Pynadath, and Paul Scerri (2001). *Adjustable Autonomy: A Response.* Intelligent Agents VII Proceedings of the International workshop on Agents, Theories, Architectures and Languages.

Yasuo Kuniyoshi (1997). *Fusing autonomy and sociability in robots*. Proceedings of the first international conference on Autonomous agents, 1997, pp. 470-471.

Lenny Foner (1997). *What's an Agent, Anyway? A Sociological Case Study*. MIT Media Lab.

Charles E. Billings (1997). *Issues Concerning Human-Centered Intelligent Systems: What's "human-centered" and what's the problem?* Plenary talk at NSF Workshop on Human-Centered Systems: Information, Interactivity, And Intelligence (HCS), February 17-19, 1997, Crystal Gateway Marriott Hotel, Arlington, VA.

Brian Scassellati (2000). *Theory of Mind for a Humanoid Robot.* The first IEEE/RSJ International Conference on Humanoid Robotics, September 2000.

Cynthia Breazeal and Brian Scassellati (1999). *How to Build Robots that Make Friends and Influence People*. Presented at the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-99), Kyongju, Korea.

Bruce Blumberg (1996). *Old Tricks, New Dogs: Ethology and Interactive Creatures.* Ph.D. thesis, MIT, chapters 1 and 2.

Justine Cassell and Hannes Vilhjálmsson (1999). *Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous.* Autonomous Agents and Multi-Agent Systems 2(1), pp. 45-64.