

# 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

## 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. Also appeared in J. Bradshaw (Ed.), (1997). *Software agents*. Menlo Park, CA and Cambridge, MA: AAI Press/The MIT Press (paper and book chapter, 6 pages)

“One of the first problems to face is that of the person's feeling of control. An important psychological aspect of people's comfort with their activities--all of their activities, from social relations, to jobs, to their interaction with technology--is the feeling of control they have over these activities and their personal lives. It's bad enough when people are intimidated by their home appliances: what will happen when automatic systems select the articles they should read, determine the importance and priority of their daily mail, and automatically answer mail, send messages, and schedule appointments? It is essential that people feel in control of their lives and surroundings, and that when automata do tasks for them, that they are comfortable with the actions, in part through a feeling of understanding, in part through confidence in the systems.”

<http://www.jnd.org/dn.mss/agents.html>

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 University, advised by Nass and Reeves (dissertation chapter, 10 pages)

“This chapter has highlighted four distinct literatures that inform the study of social responses to computer-based representations of agency. The relevance of sources of messages in general, and of self- vs. other-evaluation in particular has been explored in the context of research in Communication Social Psychology and Sociology. The perception of technologies as autonomous sources has been discussed with reference to work in both these fields and in Human-Computer Interaction (HCI) and Artificial Intelligence (AI). Other work in these fields also provided insight into the use of computers to represent human agency across a variety of different tasks and situations in an effort to create 'believable agents.' Two different classification schemes for examining

believability were presented, one that entails the belief that an entity is actually human (anthropomorphism), and one that is limited to the application of particular human characteristics to a non-human entity (ethopoeia). Finally, the relationship between conversational situations as examined in the field of Psycholinguistics and the quest for making believable computer-based representations of human-like entities was considered in light of some recent HCI and AI research projects.”

<http://www.cyborganic.com/People/jonathan/Academia/Dissertation/theory1.html>

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 (paper, 9 pages)

“We propose that the area of domestic robots is not only a suitable but also challenging field of Human-Computer Interaction, which contains its own specific research problems. The main problem statements in HCI of course remain the same, but there are additional problems that the research needs to address, e.g. the dynamic environment, object and context recognition, HCI for autonomous agents in a physical environment, just to mention a few.”

[http://www.nada.kth.se/~larsoe/AMS/Artiklar/CHI99/chi\\_ver4\\_hh.HTML](http://www.nada.kth.se/~larsoe/AMS/Artiklar/CHI99/chi_ver4_hh.HTML)

Valentino Braitenberg (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge MA: The MIT Press (book, 155 pages, get overview)

e.g., <http://www.santafe.edu/~shalizi/reviews/vehicles/>  
<http://www.santafe.edu/~shalizi/reviews/vehicles/>

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 (paper, 21 pages)

<http://orawww.cs.herts.ac.uk/~comqkd/papers.html>

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 (paper, 39 pages)

<http://orawww.cs.herts.ac.uk/~comqkd/papers.html>

Cynthia Breazeal (1999). *Robot in Society: Friend or Appliance?* In Agents99 Workshop on Emotion-Based Agent Architectures, Seattle, WA, pp. 18-26 (paper, 9 pages)

<http://www.ai.mit.edu/projects/sociable/publications.html>

David Stork (ed.) (1997). *HAL's legacy: 2001's computer as dream and reality*. Cambridge MA: The MIT Press (book, 384 pages, chapters 1,2 and 9)

<http://mitpress.mit.edu/e-books/Hal/>

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 (paper, 2 pages)

“Attempts to generate anthropomorphic responses to computers have been based on complex, agent-based interfaces. This study provides experimental evidence that minimal social cues can induce computer-literate individuals to use social rules-praise of others is more valid than praise of self, praise of others is friendlier than praise of self, and criticism of others is less friendly than criticism of self-to evaluate the performance of computers. We also demonstrate that different voices are treated as distinct agents.”

<http://www.acm.org/pubs/citations/proceedings/chi/259964/p111-nass/>

<http://www.cyborganic.com/People/jonathan/Academia/Papers/ Acrobat/interchi-93.pdf>

<http://www.cyborganic.com/People/jonathan/Academia/Papers/Web/interchi-93.html>

Kerstin Dautenhahn (2000). *Socially Intelligent Agents and The Primate Social Brain - Towards a Science of Social Minds*. Proceedings of AAI Fall Symposium *Socially Intelligent Agents - The Human in the Loop*, AAAI Press, Technical Report FS-00-04, pp. 35-51 (paper, 17 pages)  
<http://orawww.cs.herts.ac.uk/~comqkd/papers.html>

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 (book chapter, 40 pages)

"This paper is a good overview on my research agenda. The paper discusses issues of embodiment and social interaction both on the level of an individual agent as well as on the level of society. The paper addresses biological, robotic and virtual agents. Robotic experiments on imitation and a robot-human interaction are described, as well as the AURORA project."

<http://orawww.cs.herts.ac.uk/~comqkd/papers.html>

<http://link.springer.de/link/service/series/0558/bibs/1562/15620102.htm>

<http://link.springer.de/link/service/series/0558/tocs/t1562.htm#toc1562>

Robert D. Putnam (2000). *Bowling alone: The Collapse and Revival of American Community*. New York, NY: Simon and Schuster (book, 541 pages, selected chapters)

<http://www.bowlingalone.com/>, or probably rather

Robert D. Putnam (1995). *Bowling Alone: America's Declining Social Capital*. *Journal of Democracy* 6:1, January 1995, pp. 65-78 (paper, 13 pages)

[http://www.press.jhu.edu/demo/journal\\_of\\_democracy/v006/6.1putnam.html](http://www.press.jhu.edu/demo/journal_of_democracy/v006/6.1putnam.html)

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 (book, 501 pages, selected chapters)

Chapter 4: Pp. 53-68: "Computing machinery and intelligence" (Turing)

Chapter 5: Pp. 69-95: "The Turing Test: A Coffeehouse Conversation" (Hofstadter)

Chapter 8: Pp. 109-115: "The Soul of the Mark III Beast" (Miedaner)

Chapter 10: Pp. 124-146 "Selfish Genes and Selfish Memes" (Dawkins)

Chapter 11: Pp. 149-201: "Prelude... Ant Fugue" (Hofstadter)

Chapter 13: Pp. 217-231 "Where Am I?" (Dennett)

Chapter 18: Pp. 287-295 "How Trurl's Own Perfection Led to No Good" (Lem)

Chapter 22: Pp. 351-382 "Minds, Brains, and Programs" (Searle)

e.g., <http://www.california.com/~rpcman/TMI.HTM>

Byron Reeves and Clifford Nass (1996). *The Media Equation*. Stanford, CA: Cambridge University Press, selected chapters (book, 317 pages, selected chapters)

e.g., <http://www.thenetnet.com/schmeb/schmeb15.html>

Anne Foerst (1995). *The Courage to Doubt: How to Build Android Robots as a Theologian*. Talk, presented at Harvard Divinity School, November 27, 1995 (talk, 7 pages)

"The title of this talk I have chosen in accordance with the central expression in the theology of Paul Tillich: The Courage to Be. And I will explain the meaning of this Tillichian expression and its importance for any dialogue between supporters of Artificial Intelligence (AI) and its opponents in four steps:

- I will describe a project at MIT as one example for AI-projects which create many hopes, but also many fears and, therefore, opposition.
- I will outline the underlying assumptions and hopes of this project.
- I will describe the arguments of the opponents of this and other similar projects and will argue why these arguments necessarily have to fail.

- I will briefly introduce some ideas of Tillich on polarities and ambiguities of human life and will show to what extent this theological concept can establish a dialogue in which both sides, AI and theology, can enrich each other.”

<http://www.ai.mit.edu/people/annef/courage/brownbag/brownbag.html>

<http://www.ai.mit.edu/people/annef/annef.html>

<http://www.nytimes.com/2000/11/07/science/07FOER.html>

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 (book, 300 pages, selected chapters)

<http://www.amazon.com/exec/obidos/ISBN=0716704633/>

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 [Reprinted in *CACM* 26(1): 23-28 (1983)] (paper, 10 pages)

“Eliza was the name of a family of programs that attempted to conduct conversations with humans...”

<http://www.acm.org/pubs/articles/journals/cacm/1983-26-1/p23-weizenbaum/p23-weizenbaum.pdf>

<http://acf5.nyu.edu/~mm64/x52.9265/january1966.html>

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

“Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what **beliefs** that agent ought to have, given its place in the world and its purpose. Then you figure out what **desires** it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its **goals** in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in most instances yield a decision about what the agent ought to do; that is what you predict the agent *will* do.” (p. 17)

[http://www.amazon.com/exec/obidos/ASIN/0262540533/qid=991267445/sr=1-1/ref=sc\\_b\\_1/104-6196117-5355949](http://www.amazon.com/exec/obidos/ASIN/0262540533/qid=991267445/sr=1-1/ref=sc_b_1/104-6196117-5355949)

<http://www.magma.ca/~mrw/agents/what-intentional-stance.html>

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

“Our most powerful 21st-century technologies - robotics, genetic engineering, and nanotech - are threatening to make humans an endangered species.”

<http://www.wired.com/wired/archive/8.04/joy.html>

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

“Computers in the 1990's can communicate with people through a fairly high bandwidth (video, audio, force feedback, etc.) Unfortunately people communicate with computers today use a very limited bandwidth, usually involving typing or using their mouse, but not much more. What will the world of computing be like in the next ten years? Sun has a vision of the merging of voice, video conferencing and shared work spaces. Sun's new Movie "Starfire" deals with a new high-productivity interface. This second generation interface will enable people to interact with their systems, their information spaces and with each other in a straightforward manner.”

A demo tape made by Sun Microsystems showing one vision of the office of the future. They took great liberties oversimplifying some thorny technological issues that must be solved before high-tech office environments like the one shown can be achieved. The short video shows an office worker using an interactive display desk to teleconference and telework, edit documents, spy on employees (!), prepare a presentation, etc.

<http://www.asktog.com/starfire/starfireHome.html>

Erik Brynjolfsson and Michael Smith (2000). *The Great Equalizer? Customer Choice Behavior at Internet Shopbots*. Unpublished paper (paper, 50 pages)

“Our research empirically analyzes consumer behavior at Internet shopbots— sites that allow consumers to make “one-click” price comparisons for product offerings from multiple retailers. By allowing researchers to observe exactly what information the consumer is shown and their search behavior in response to this

information, shopbot data has unique strengths for analyzing consumer behavior. Furthermore, the method in which the data is displayed to consumers lends itself to a utility-based evaluation process, consistent with econometric analysis techniques. While price is an important determinant of customer choice, we find that, even among shopbot consumers, branded retailers and retailers a consumer visited previously hold significant price advantages in head-to-head price comparisons. We also find that these models accurately predict consumer behavior out of sample, suggesting that our analyses effectively capture relevant aspects of consumer choice processes and can form a useful basis for understanding consumer behavior and leveraging this understanding to strategic advantage.”

<http://ecommerce.mit.edu/papers/tge/tge.pdf>

## 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 (paper, 6 pages)

"While the tone of this paper is informal and tongue-in-cheek, we believe we raise two important issues in robotics and multi-modal interface research; namely, how crucial integration of multiple modes of communication are for adjustable autonomy, which in turn is crucial for having dinner with R2D2. Furthermore, we discuss how our multi-modal interface to autonomous robots addresses these issues by tracking goals, allowing for both natural and mechanical modes of input, and how our robotic system adjusts itself to ensure that goals are achieved, despite interruptions."

<ftp://ftp.aic.nrl.navy.mil/pub/papers/2000/AIC-00-001.pdf>

<http://www.aic.nrl.navy.mil/~dennisp/bibliography.html>

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 (paper, 12 pages)

<http://link.springer.de/link/service/series/0558/bibs/1792/17920273.htm>

[http://www.springer.co.uk/com\\_pubs/ct\\_virtin.htm](http://www.springer.co.uk/com_pubs/ct_virtin.htm)

Michael Mogensen (2001). *Dependent Autonomy and Transparent Automaton?* In Lars Qvortrup (ed.) *Virtual Interaction: Interaction in/with Virtual Inhabited 3D Worlds*, New York, NY: Springer (book chapter, 17 pages)

<http://www.intermedia.auc.dk/staging/html/publications/publications.html>

[http://www.intermedia.auc.dk/staging/pdf/07\\_MM.pdf](http://www.intermedia.auc.dk/staging/pdf/07_MM.pdf)

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 (paper, 7 pages)

"We are designing and implementing a multi-modal interface to an autonomous robot. For this interface, we have elected to use natural language and gesture. Gestures can be either natural gestures perceived by a vision system installed on the robot, or they can be made by using a stylus on a Personal Digital Assistant. In this paper we describe how we are attempting to provide a seamless integration of the various modes of input to provide a multi-modal interface that humans can manipulate as they desire. The interface will allow the user to choose whatever mode or combination of modes seems appropriate for interactions with the robot. The human user, therefore, does not have to be limited to any one mode of interaction, but can freely choose whatever mode is most comfortable or natural."

<ftp://ftp.aic.nrl.navy.mil/pub/papers/2000/AIC-00-003.pdf>

<http://www.aic.nrl.navy.mil/~dennisp/bibliography.html>

Eric Horvitz (1999). *Principles of Mixed-Initiative User Interfaces*. ACM CHI'99 Proceedings, pp. 159-166 (paper, 8 pages)

"Recent debate has centered on the relative promise of focusing user-interface research on developing new metaphors and tools that enhance users' abilities to directly manipulate objects versus directing effort toward developing interface agents that provide automation. In this paper, we review principles that show promise for allowing engineers to enhance human-computer interaction through an elegant coupling of automated services with direct manipulation."

<http://www.acm.org/pubs/citations/proceedings/chi/302979/p159-horvitz/>

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 (paper, 7 pages)

<http://www.acm.org/pubs/citations/proceedings/uist/238218/p33-shneiderman/>

<http://www.cs.umd.edu/hcil/members/bshneiderman/umlpapers/articles.html>

Marc Mersiol, Ayda Saidane (2000). *A Tool to Support Function Allocation*. Proceedings of Safety and Usability Concerns in Aeronautics, SUCA 2000 (paper, 5 pages)

"The scope of this position paper is to present a tool able to help designers to allocate functions. Function allocation refers to the attribution of functions between agents (humans and machines) in a sociotechnical system early in the design process. We present existing function allocation methods and discuss two of their main drawbacks. We propose directions for overcoming these limits and describe a tool supporting function allocation decisions."

<http://lis.univ-tlse1.fr/~palanque/WSSUCA2000/suca-Mersiol.pdf>

<http://lis.univ-tlse1.fr/~palanque/SUCA2000.htm>

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 (paper, 22 pages)

[http://ic-www.arc.nasa.gov/ic/projects/Executive/papers/mars\\_adj\\_auton98.pdf](http://ic-www.arc.nasa.gov/ic/projects/Executive/papers/mars_adj_auton98.pdf)

Alan Wexelblat and Pattie Maes (1997). *Issues for Software Agent UI*. Unpublished paper (paper, 18 pages)

"Agent user interfaces pose a number of special challenges for interface designers. These challenges can be formulated as a series of issues which must be addressed: understanding, trust, control, distraction, and personification. We examine each of these in turn and draw recommendations for designers in dealing with each of the issues as well as for the overall design of an agent interface based on our experiences with building such systems."

<http://wex.www.media.mit.edu/people/wex/agent-ui-paper/agent-ui.htm>

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 (article, 20 pages)

"Ben Shneiderman is a long-time proponent of direct manipulation for user interfaces. Direct manipulation affords the user control and predictability in their interfaces. Pattie Maes believes direct manipulation will have to give way to some form of delegation—namely software agents. Should users give up complete control of their interaction with interfaces? Will users want to risk depending on "agents" that learn their likes and dislikes and act on a user's behalf? Ben and Pattie debated these issues and more at both IUI 97 (Intelligent User Interfaces conference - January 6-9, 1997) and again at CHI 97 in Atlanta (March 22-27, 1997). Read on and decide for yourself where the future of interfaces should be headed—and why."

<http://www.it-uni.sdu.dk/mmp/Library/ShneidermanMaes97.pdf>

## 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 (paper, 7 pages)

"Unmanned vehicles, such as mobile robots, must exhibit adjustable autonomy. They must be able to be self-sufficient when the situation warrants; however, as they interact with each other and with humans, they must exhibit an ability to dynamically adjust their independence or dependence as co-operative agents attempting to achieve some goal. This is what we mean by adjustable autonomy. We have been investigating various modes of communication that enhance a robot's capability to work interactively with other robots and with humans. Specifically, we have been investigating how natural language and gesture can provide a user-friendly interface to mobile robots. We have extended this initial work to include semantic and pragmatic procedures that allow humans and robots to act co-operatively, based on whether or not the goal has been achieved. The various agents involved in achieving the goals are each aware of their own and others' goals and what goals have been stated or accomplished so that eventually any member of the group, be it a robot or a human, if necessary, can interact with the other members to achieve the stated goals of a mission."

<ftp://ftp.aic.nrl.navy.mil/pub/papers/2000/AIC-00-002.pdf>

<http://www.aic.nrl.navy.mil/~dennisp/bibliography.html>

Phoebe Sengers, Simon Penny, and Jeffrey Smith (2000). *Traces: Semi-Autonomous Avatars*. Unpublished paper (paper, 5 pages)

"This paper describes work on Traces, a Virtual Reality system which allows full-body, physical interaction with a variety of avatars. We argue that avatars should be thought of, not as simple representations of users, but on a range of autonomy levels from classical avatars through autonomous agents. We describe 3 levels of autonomy in the Traces avatars."

<http://www.cs.cmu.edu/afs/cs.cmu.edu/user/phoebe/mosaic/work/publications.html>

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 (paper, 15 pages)

"This paper discusses the role of predictability and control in robot-human interaction. This involves the central question whether humans are good models for synthetic (social) agents."

<http://www.cogtech.org/CT99/Dautenhahn.htm>

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 (paper, 3 pages)

<http://www.isi.edu/teamcore/elvespapers.html>

<http://www.isi.edu/teamcore/papers.html>

Yasuo Kuniyoshi (1997). *Fusing autonomy and sociability in robots*. Proceedings of the first international conference on Autonomous agents, 1997, pp. 470-471 (paper, 2 pages)

<http://www.acm.org/pubs/citations/proceedings/ai/267658/p470-kuniyoshi/>

Lenny Foner (1997). *What's an Agent, Anyway? A Sociological Case Study*. MIT Media Lab (paper, 40 pages)

<http://foner.www.media.mit.edu/people/foner/Julia/Julia.html>

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 (paper of talk)

"Humans are responsible for outcomes in human-machine systems" (...) "Automation that is strong, silent, and hard to direct is \*not\* a team player". (...) "If autonomous behavior is unexpected by a human operator, it is often perceived as "animate"; the machine appears to have a "mind of its own". The human must decide whether the perceived behavior is appropriate, or whether it represents a failure of the machine component of the system. This decision can be rather difficult." (...) "I suggest that machines that are compliant with our demands, communicative regarding their processes, and cooperative in our endeavors can indeed be team players - and team play is at the heart of a human-centered intelligent system."

<http://www.ifp.uiuc.edu/nsfhcs/talks/billings.html>

Brian Scassellati (2000). *Theory of Mind for a Humanoid Robot*. The first IEEE/RSJ International Conference on Humanoid Robotics, September 2000 (paper, 12 pages)

<http://www.ai.mit.edu/~scasz/papers/Humanoids2000-tom.pdf>

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 (paper, 6 pages)

"In order to interact socially with a human, a robot must convey intentionality, that is, the human must believe that the robot has beliefs, desires, and intentions. We have constructed a robot which exploits natural human social tendencies to convey intentionality through motor actions and facial expressions. We present results on the integration of perception, attention, motivation, behavior, and motor systems which allow the robot to engage in infant-like interactions with a human caregiver."

<http://www.ai.mit.edu/~scasz/papers/Breazeal-Scasz-IROS99.pdf>

Bruce Blumberg (1996). *Old Tricks, New Dogs: Ethology and Interactive Creatures*. Ph.D. thesis, MIT, chapters 1 and 2 (thesis chapters, 16 pages)

"This thesis seeks to address the problem of building things with behavior and character. By things we mean autonomous animated creatures or intelligent physical devices. By behavior we mean that they display the rich level of behavior found in animals. By character we mean that the viewer should "know" what they are "feeling" and what they are likely to do next."

"In this chapter we have argued that some level of autonomy is desirable in many, if not all, interactive characters. However, we also stressed the point that autonomy is not an all or nothing thing, but rather differing degrees of autonomy may be desired depending on the application. Our point in discussing what we saw as necessary components for creating the "illusion of life", was to stress that "life-like" means more than simply pos-sessing autonomy. Indeed, an important characteristic of being "life-like" is the ability to convey intentionality and, possibly conflicting motivational states, through movement and the quality of that movement. Finally, we described a number of practical applications for these kinds of creatures, and discussed the broader applicability of this work."

[http://characters.www.media.mit.edu/groups/characters/thesis/blumberg\\_phd.pdf](http://characters.www.media.mit.edu/groups/characters/thesis/blumberg_phd.pdf)

Justine Cassell and Hannes Vilhjálmsón (1999). *Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous*. Autonomous Agents and Multi-Agent Systems 2(1), pp. 45-64 (paper, 21 pages)

"Modeling and animation of gestures is crucial for the credibility and effectiveness of the virtual interaction in chat. By treating the avatar as a communicative agent, we propose a method to automate the animation of important communicative behavior, deriving from work in conversation and discourse theory."

[http://gn.www.media.mit.edu/groups/gn/publications/agents\\_journal99.pdf](http://gn.www.media.mit.edu/groups/gn/publications/agents_journal99.pdf)