SOCIALLY INTELLIGENT AGENTS

Creating Relationships with Computers and Robots

Edited by

Kerstin Dautenhahn

Alan H. Bond

Lola Cañamero

Bruce Edmonds

Kluwer Academic Publishers

SOCIALLY INTELLIGENT AGENTS

Creating Relationships with Computers and Robots

MULTIAGENT SYSTEMS, ARTIFICIAL SOCIETIES, AND SIMULATED ORGANIZATIONS

International Book Series

Series Editor: Gerhard Weiss
Technische Universität München

Books in the Series:

CONFLICTING AGENTS: ***Conflict Management in Multi-Agent Systems***, edited by Catherine Tessier, Laurent Chaudron and Heinz-Jürgen Müller, ISBN: 0-7923-7210-7

SOCIAL ORDER IN MULTIAGENT SYSTEMS, edited by Rosaria Conte and Chrysanthos Dellarocas, ISBN: 0-7923-7450-9

CONCEPTUAL MODELLING OF MULTI-AGENT SYSTEMS: ***The CoMoMAS Engineering Environment***, by Norbert Glaser, ISBN: 1-4020-7061-6

SOCIALLY INTELLIGENT AGENTS

Creating Relationships with Computers and Robots

Edited by

Kerstin Dautenhahn
University of Hertfordshire

Alan H. Bond
California Institute of Technology

Lola Cañamero
University of Hertfordshire

Bruce Edmonds
Manchester Metropolitan University

KLUWER ACADEMIC PUBLISHERS
Boston / Dordrecht / London

eBook ISBN: 0-306-47373-9
Print ISBN: 1-4020-7057-8

Created in the United States of America

Visit Kluwer Online at: http://www.kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

Contents

Contributing Authors

Aude Billard
Computer Science Department, University of Southern California, HNB, 3641 Wyatt Way, Los Angeles 90089, USA. billard@usc.edu

Katharine Blocher
Formerly of Massachusetts Institute of Technology, Media Laboratory, 4615 Huron Ave., San Diego, CA 92117, USA. kblocher@alum.mit.edu

Alan H. Bond
California Institute of Technology, Department of Computer Science, Mailstop 256-80, Pasadena, CA 91125, USA. bond@cs.caltech.edu

Cynthia Breazeal
The Media Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Ave., NE18-5FL, Cambridge, MA 02139-4307, USA.
cynthiab@media.mit.edu

Paul Brna
Computer Based Learning Unit, University of Leeds, Leeds LS2 9JT, United Kingdom. P.Brna@cbl.leeds.ac.uk

Lola Cañamero
Adaptive Systems Research Group, Department of Computer Science, University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, United Kingdom. L.Canamero@herts.ac.uk

Edmund Chattoe
University of Oxford, Department of Sociology, Littlegate House, St Ebbes, Oxford, OX1 1PT, United Kingdom.
edmund.chattoe@sociology.oxford.ac.uk

Cristina Conati
Department of Computer Science, University of British Columbia, 2366 Main Mall, Vancouver, B.C. Canada V6T 1Z4. conati@cs.ubc.ca

Bridget Cooper
Computer Based Learning Unit, University of Leeds, Leeds LS2 9JT, United Kingdom. B.L.Cooper@cbl.leeds.ac.uk

Kerstin Dautenhahn
Adaptive Systems Research Group, Department of Computer Science, University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, United Kingdom. K.Dautenhahn@herts.ac.uk

Berardina Nadja De Carolis
Intelligent Interfaces, Department of Informatics, University of Bari, Via Orabona 4, 70126 Bari, Italy. decarolis@di.uniba.it

Fiorella de Rosis
Intelligent Interfaces, Department of Informatics, University of Bari, Via Orabona 4, 70126 Bari, Italy. derosis@di.uniba.it

Paul Dickerson
University of Surrey Roehampton, School of Psychology and Counselling, Whitelands College, West Hill, London, SW15 3SN, United Kingdom.
p.dickerson@roehampton.ac.uk

Allison Druin
Institute for Advanced Computer Studies, University of Maryland, College Park, MD 742, USA. allisond@umiacs.umd.edu

Bruce Edmonds
Centre for Policy Modelling, Manchester Metropolitan University, Aytoun Building, Aytoun Street, Manchester, M1 3GH, United Kingdom.
b.edmonds@mmu.ac.uk

Peyman Faratin
Center for Coordination Science, MIT Sloan School of Management, NE20-336, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
peyman@mit.edu

Jonathan Gratch
USC Institute for Creative Technologies, 13274 Fiji Way, Suite 600, Marina del Rey, CA 90292, USA. gratch@ict.usc.edu

James A. Hendler
Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA. hendler@cs.umd.edu

Eva Hudlicka
Psychometrix Associates, Inc., 1805 Azalea Drive, Blacksburg, VA 24060, USA. hudlicka@acm.org

Katherine Isbister
Finali Corporation, 3001 19th Street, 2nd floor, San Francisco, CA 94110, USA. kath@cyborganic.net

Maria Klawe
Department of Computer Science, University of British Columbia, 2366 Main Mall, Vancouver, B.C. Canada V6T 1Z4. klawe@interchg.ubc.ca

Hideki Kozima
Social Interaction Group, Keihanna Human Info-Communication Research Center, Communications Research Laboratory, 2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan. xkozima@crl.go.jp

Hidekazu Kubota
Faculty of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. kubota@kc.t.u-tokyo.ac.jp

Jarmo Laaksolahti
Swedish Institute of Computer Science (SICS), Box 1263,
SE-164 29 Kista, Sweden. jarmo@sics.se

Peter Lönnqvist
Department of Computer and Systems Sciences, Stockholm University and Royal Institute of Technology, Stockholm, Sweden. peterl@dsv.su.se

Isabel Machado
Instituto de Engenharia de Sistemas e Computadores (INESC), Rua Alves Redol 9, 1100 Lisboa, Portugal. Isabel.Machado@inesc.pt

Stacy Marsella
USC Information Sciences Institute, 4676 Admiralty Way, Suite 1001, Marina del Rey, CA 90292, USA. marsella@isi.edu

Michael Mateas
Computer Science Department, Carnegie-Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. michaelm@cs.cmu.edu

Helen McBreen
Centre for Communication Interface Research, Department of Electronics and Electrical Engineering, University of Edinburgh, 80 South Bridge, EH1 1HN, United Kingdom. Helen.McBreen@ccir.ed.ac.uk

François Michaud
Department of Electrical Engineering and Computer Engineering, Université de Sherbrooke, 2500 boul. Université, Sherbrooke (Québec) J1K 2R1, Canada. francois.michaud@courrier.usherb.ca

Jaime Montemayor
Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA. monte@cs.umd.edu

Scott Moss
Centre for Policy Modelling, Manchester Metropolitan University, Aytoun Building, Aytoun Street, Manchester, M1 3GH, United Kingdom. s.moss@mmu.ac.uk

Toyoaki Nishida
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.
nishida@kc.t.u-tokyo.ac.jp

Bernard Ogden
Adaptive Systems Research Group, Department of Computer Science, University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, United Kingdom. bernard@aurora-project.com

Ana Paiva
Instituto de Engenharia de Sistemas e Computadores (INESC), Rua Alves Redol 9, 1100 Lisboa, Portugal. Ana.Paiva@inesc.pt

Valery A. Petrushin
Center for Strategic Technology Reasearch, Accenture, 3773 Willow Road, Northbrook, IL 60062, USA. petr@cstar.accenture.com

Per Persson
Swedish Institute of Computer Science (SICS), Box 1263,
SE-164 29 Kista, Sweden. perp@sics.se

Rosalind W. Picard
Massachusetts Institute of Technology, Media Laboratory, 20 Ames Street, Cambridge, MA 02139, USA. picard@media.mit.edu

John P. Pinto
Formerly of Interval Research Corporation. johnppinto@yahoo.com

Sebastiano Pizzutilo
Intelligent Interfaces, Department of Informatics, University of Bari, Via Orabona 4, 70126 Bari, Italy. pizzutilo@di.uniba.it

David V. Pynadath
Information Sciences Institute, University of Southern California, 4676 Admiralty Way, Marina del Rey, CA 90292, USA. pynadath@isi.edu

John Rae
University of Surrey Roehampton, School of Psychology and Counselling, Whitelands College, West Hill, London, SW15 3SN, United Kingdom.
j.rae@roehampton.ac.uk

Krisnawan Rahardja
Formerly of Interval Research Corporation. rahardkk@yahoo.com

Juan A. Rodríguez-Aguilar
iSOCO Barcelona, Alcalde Barnils, 64-68 Edificio Testa - bl. A, 08190 Sant Cugat Del Valles, Spain. Formerly of IIIA, Spanish Scientific
Research Council (CSIC), Spain. jar@isoco.com

Juliette Rouchier
GREQAM (CNRS), 2 Rue de la Charite, 13002 Marseille, France.
rouchier@ehess.cnrs-mrs.fr

Mark Scheeff
Formerly of Interval Research Corporation. mark@markscheeff.com

Scott Sona Snibbe
Formerly of Interval Research Corporation. scott@snibbe.com

Carles Sierra
Institut d'Investigació en Intel.ligència Artificial (IIIA), Spanish Scientific
Research Council (CSIC), Campus de la UAB, 08193 Bellaterra, Spain.
sierra@iiia.csic.es

Andrew Stern
www.interactivestory.net, andrew@interactivestory.net

Penny Stribling
University of Surrey Roehampton, School of Psychology and Counselling,
Whitelands College, West Hill, London, SW15 3SN, United Kingdom.
P.Stribling@btinternet.com

Milind Tambe
Information Sciences Institute, University of Southern California,
4676 Admiralty Way, Marina del Rey, CA 90292, USA. tambe@isi.edu

Nell Tenhaaf
Department of Visual Arts, 232 Centre for Fine Arts, York University, 4700 Keele Street, Toronto, Ontario, Canada, M3J 1P3.
tenhaaf@yorku.ca

Catherine Théberge-Turmel
Department of Electrical Engineering and Computer Engineering, Université de Sherbrooke, 2500 boul. Université, Sherbrooke (Québec) J1K 2R1, Canada.
catherine.t@hermes.usherb.ca

Robert Tow
AT & T Labs, 75 Willow Road, Menlo Park, CA 94025, USA.
rtow@attlabs.att.com

Iain Werry
Department of Cybernetics, University of Reading, Whiteknights,
PO Box 225, Reading, Berks RG6 6AY, United Kingdom.
Iain@aurora-project.com

R. Michael Young
Department of Computer Science, Box 8206, College of Engineering, North Carolina State University, Raleigh, NC 27695, USA.
young@csc.ncsu.edu

Chapter 1

SOCIALLY INTELLIGENT AGENTS

Creating Relationships with Computers and Robots

Kerstin Dautenhahn[1], Alan Bond[2], Lola Cañamero[1], and Bruce Edmonds[3]
[1]*University of Hertfordshire,* [2]*California Institute of Technology,* [3]*Manchester Metropolitan University*

Abstract This introduction explains the motivation to edit this book and provides an overview of the chapters included in this book. Main themes and common threads that can be found across different chapters are identified that might help the reader in navigating the book.

1. Background: Why this book?

The field of Socially Intelligent Agents (SIA) is by many perceived as a growing and increasingly important research area that comprises very active research activities and strongly interdisciplinary approaches. The field of Socially Intelligent Agents is characterized by agent systems that show human-style social intelligence [5]. Humans live in individualized societies where group members know each other, so do other animal species, cf. figure 1.1. Although overlap exists, SIA systems are different from multi-agent systems that a) are often only loosely related to human social intelligence, or use very different models from the animal world, e.g. self-organization in social insect societies, or b) might strongly focus on the engineering and optimization aspects of the agent approach to software engineering.

In the past, two AAAI Fall Symposia were organized on the topic of Socially Intelligent Agents, in 1997 and 2000. Both symposia attracted a large number of participants. The first symposium gave a general overview on the spectrum of research in the field, and in the years following this event a variety of publications (special journal issues and books) resulted from it[1]. Also, a number of related symposia and workshops were subsequently organized[2]. Unlike the 1997 symposium, the 2000 symposium specifically addressed the issue of Socially Intelligent Agents - The Human in the Loop. A special issue

Figure 1.1. Elephants are socially intelligent biological agents that live in family groups with strong, long-lasting social bonds. Much research into socially intelligent artifacts is inspired by animal (including human) social intelligence.

of *IEEE Systems, Man and Cybernetics, Part A* emerged from this symposium which provides an in depth treatment of a few research approaches in that area[3]. Unlike the special journal issue, this book has a radically different nature: it is intended to be the first definitive collection of current work in the rapidly growing field of Socially Intelligent Agents, providing a useful and timely reference for computer scientists, web programmers and designers, computer users, and researchers interested in the issue of how humans relate to computers and robots, and how these agents in return can relate to them. Each of the 32 chapters is, compared to a journal article, relatively short and compact, focusing on the main theoretical and practical issues involved in the work. Each chapter gives references to other publications that can provide the reader with further detailed information.

In the area of software and intelligent agents many other publications are available, e.g. [1], [9], [6], proceedings of the *Autonomous Agents* and other conferences, just to name a few. However, none of them provide a state-of-the-art *reference book* on Socially Intelligent Agents with an interdisciplinary approach including both software and robotic agents.

Despite many publications that either a) specialize in particular issues relevant to Socially Intelligent Agents (e.g. robots, emotions, conversational skills, narrative, social learning and imitation etc., cf. [12], [10], [3], [7], [2], [11], [4]), or b) present a small number of in-depth discussions of particular research projects (published in journal issues mentioned above), the field of Socially Intelligent Agents is missing a state-of-the-art collection that can provide an overview and reference book. More and more researchers and PhD students are interested in learning about and participating in SIA research, but at present the only way to learn about the field is to go through and select among a large number of widely ‘distributed’ and often difficult to access publications, i.e. journal issues, books, conference and workshop proceedings etc. Our motivation to edit this book was therefore based on the belief that there is a strong demand for a book that can be used by students, researchers and anybody interested in learning about Socially Intelligent Agents. The main strength of the book is the breadth of research topics presented and the references given at the end of each chapter, so that researchers who want to work in that field are given pointers to literature and other important work not included in the book.

The book presents a coherent and structured presentation of state-of-the-art in the field. It does not require the reader to possess any specialist knowledge and is suitable for any student / researcher with a general background in Computer Science and/or Artificial Intelligence or related fields (e.g. Cognitive Science, Cybernetics, Adaptive Behavior, Artificial Life etc.). Also, at present the growing field of Socially Intelligent Agents has no core text that can be used in university courses. This book fills this gap and might be used in different courses for postgraduate studies, and as research material for PhD students, e.g. for studies in Applied Artificial Intelligence, Intelligent and Autonomous Agents, Adaptive Systems, Human-Computer Interaction, or Situated, Embodied AI.

2. Book Structure and Chapter Overviews

The remaining thirty-two chapters of this book are organized into two parts. The structure of the book is visually shown in figure 1.2. The first part addresses the theory, concepts and technology of Socially Intelligent Agents. The second part addresses current and potential applications of Socially Intelligent Agents. The first part of the book has twelve chapters organized in three sections covering three major themes, namely relationships between agents and humans, edited by Alan Bond, agents and emotions/personality edited by Lola Cañamero, and communities of social agents, edited by Bruce Edmonds. The second part of the book consists of twenty chapters organized in five sections covering the themes of interactive therapeutic agent systems, edited by Kerstin Dautenhahn, socially intelligent robots, edited by Lola Cañamero, interactive education and training, edited by Kerstin Dautenhahn, social agents in games and entertainment, edited by Alan Bond, and social agents in e-commerce, edited by Bruce Edmonds. The content of the sections and chapters is described in more detail below.

Note, that thematically we have strong overlaps between all chapters in this book, the division into thematic sections is mainly of practical nature. This

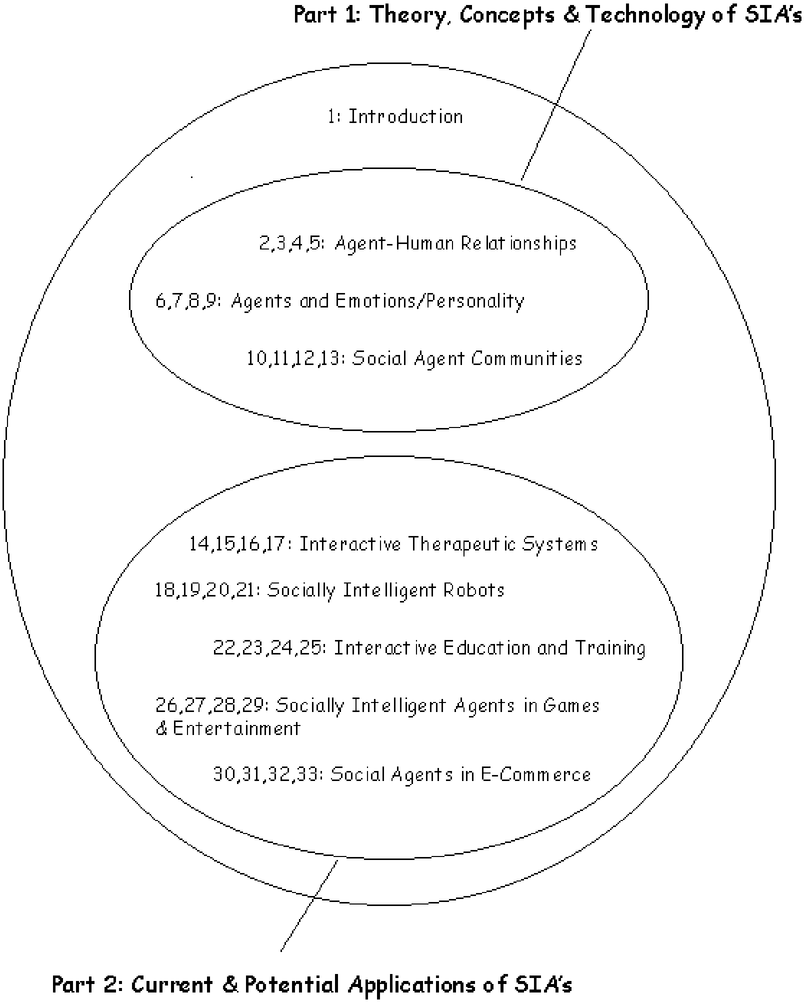

Part 1: Theory, Concepts & Technology of SIA's
1: Introduction
2,3,4,5: Agent-Human Relationships
6,7,8,9: Agents and Emotions/Personality
10,11,12,13: Social Agent Communities
14,15,16,17: Interactive Therapeutic Systems
18,19,20,21: Socially Intelligent Robots
22,23,24,25: Interactive Education and Training
26,27,28,29: Socially Intelligent Agents in Games & Entertainment
30,31,32,33: Social Agents in E-Commerce
Part 2: Current & Potential Applications of SIA's

Figure 1.2. Book structure, showing the division into two parts and eight sections. Chapter numbers are given.

introductory chapter therefore concludes by identifying a few of these thematic overlaps (section 3).

2.1 Agent-Human Relationships

This first section engages the reader in the question of what a relationship between a computer agent and a human user might be. Are relationships possible at all, and if so, what would it mean for an agent and a human to have a relationship? What theoretical bases should we use for this problem? How can we design and implement agents that engage in and maintain relationships with users? How will we be able to provide and to manage such agents?

There are a number of dimensions of analysis of this problem, such as:

- What interaction methods and protocols are efficacious?
- What kinds of information should be exchanged?
- What knowledge can be and should be shared?
- How do we model the other?
  - How should a computer agent model the human?
  - How will the human user model or think of the computer agent?
- What kinds of constraints on behavior of both partners can result, how do we represent them, communicate them, detect them, renegotiate them? and
- What are the effects, benefits and drawbacks of agent-human relationships?

Chapter 2, written by Per Persson, Jarmo Laaksolahti, and Peter Lönnqvist presents a social psychological view of agent-human relationships, drawing on their backgrounds in cultural studies and film. They observe that users adopt an intentional instead of mechanical attitude in understanding socially intelligent agents, pointing out the active role of the human mind in constructing a meaningful reality. According to their constructivist approach, socially intelligent agents must be meaningful, consistent and coherent to the user. In order to characterize this mentality, the authors draw upon a comprehensive background including folk psychology and trait theory. They advocate the use of folk theories of intelligence in agent design, however this will be idiosyncratic to the user and their particular culture.

In chapter 3, Alan Bond discusses an implemented computer model of a socially intelligent agent, and its dynamics of relationships between agents and between humans and agents. He establishes two main properties of his model which he suggests are necessary for agent-human relationships. The first is voluntary action and engagement: agents, and humans, must act voluntarily and autonomously. The second is mutual control: in a relationship humans and agents must exert some control over each other. The conciliation of these two principles is demonstrated by his model, since agents voluntarily enter into mutually controlling regimes.

Bruce Edmonds presents in chapter 4 a very interesting idea that might be usable for creating socially intelligent agents. He suggests that agents be created using a developmental loop including the human user. The idea is for the agent to develop an identity which is intimately suited to interaction with that particular human. This, according to the author may be the only way to achieve the quality of relationship needed. In order to understand such a process, the author draws upon current ideas of the human self and its ontogenetic formation. He articulates a model of the construction of a self by an agent, in interaction with users.

In chapter 5, Katherine Isbister discusses the use of nonverbal social cues in social relationships. Spatial proximity, orientation and posture can communicate social intention and relationship, such as agreement or independence among agents. Facial expressions and hand, head and body gestures can indicate attitude and emotional response such as approval or uncertainty. Spatial pointing and eye gaze can be used to indicate subjects of discussion. Timing, rhythm and emphasis contribute to prosody and the management of conversational interaction. Her practical work concerns the development of interface agents whose purpose is to facilitate human-human social interaction. She reports on her experience in two projects, a helper agent and a tour guide agent.

2.2 Agents and Emotions/Personality

Emotion is key in human social activity, and the use of computers and robots is no exception. Agents that can recognize a user's emotions, display meaningful emotional expressions, and behave in ways that are perceived as coherent, intentional, responsive, and socially/emotionally appropriate, can make important contributions towards achieving human-computer interaction that is more 'natural', believable, and enjoyable to the human partner. Endowing social artifacts with aspects of personality and emotions is relevant in a wide range of practical contexts, in particular when (human) trust and sympathetic evaluation are needed, as in education, therapy, decision making, or decision support, to name only a few.

Believability, understandability, and the problem of realism are major issues addressed in the first three chapters of this section, all of them concerned with different aspects of how to design (social) artifacts' emotional displays and behavior in a way that is adapted to, and recognizable by humans. The fourth chapter addresses the converse problem: how to build agents that are able to recognize human emotions, in this case from vocal cues.

In chapter 6, Eva Hudlicka presents the *ABAIS* adaptive user interface system, capable of recognizing and adapting to the user's affective and belief states. Based on an adaptive methodology designed to compensate for performance biases caused by users' affective states and active beliefs, ABAIS provides a generic framework for exploring a variety of user affect assessment methods and GUI adaptation strategies. The particular application discussed in this chapter is a prototype implemented and demonstrated in the context of an Air Force combat task. Focusing on traits ‘anxiety’, ‘aggressiveness’, and ‘obsessiveness’, the prototype uses a knowledge-based approach to assess and adapt to the pilot’s anxiety level by means of different task-specific compensatory strategies implemented in terms of specific GUI adaptations. One of the focal goals of this research is to increase the realism of social intelligent agents in situations where individual adaptation to the user is crucial, as in the critical application reported here.

Chapter 7, by Sebastiano Pizzutilo, Berardina De Carolis, and Fiorella De Rosis discusses how cooperative interface agents can be made more believable when endowed with a model that combines the communication traits described in the Five Factor Model of personality (e.g., ‘extroverted’ versus ‘introverted’) with some cooperation attitudes. Cooperation attitudes refer in this case to the level of help that the agent provides to the user (e.g., an overhelper agent, a literal helper agent), and the level of delegation that the user adopts towards the agent (e.g., a lazy user versus a ‘delegating-if-needed’ one). The agent implements a knowledge-based approach to reason about and select the most appropriate response in every context. The authors explain how cooperation and communication personality traits are combined in an embodied animated character (XDM-Agent) that helps users to handle electronic mail using Eudora.

In chapter 8, Lola Cañamero reports the rationale underlying the construction of Feelix, a very simple expressive robot built from commercial LEGO technology, and designed to investigate (facial) emotional expression for the sole purpose of social interaction. Departing from realism, Cañamero’s approach advocates the use of a ‘minimal’ set of expressive features that allow humans to recognize and analyze meaningful basic expressions. A clear causal pattern of emotion elicitation—in this case based on physical contact—is also necessary for humans to attribute intentionality to the robot and to make sense of its displays. Based on results of recognition tests and interaction scenarios, Cañamero then discusses different design choices and compares them with some of the guidelines that inspired the design of other expressive robots, in particular Kismet (cf. chapter 18). The chapter concludes by pointing out some of the ‘lessons learned’ about emotion from such a simple robot.

Chapter 9, by Valery Petrushin, investigates how well people and computers can recognize emotions in speech, and how to build an agent that recognizes emotions in speech signal to solve practical, real-world problems. Motivated by the goal of improving performance at telephone call centers, this research addresses the problem of detecting emotional state in telephone calls with the purpose of sorting voice mail messages or directing them to the appropriate person in the call center. An initial research phase, reported here, investigated which features of speech signal could be useful for emotion recognition, and explored different machine learning algorithms to create reliable recognizers.

This research was followed by the development of various pieces of software—among others, an agent capable of analyzing telephone quality speech and to distinguish between two emotional states—‘agitation’ and ‘calm’—with good accuracy.

2.3 Social Agent Communities

Although it has always been an important aspect of agents that they distribute computation using local reasoning, the consequences of this in terms of the increased complexity of coordination between the agents were realized more slowly. Thus, in recent years, there has been a move away from designing agents as single units towards only studying and implementing them as whole societies. For the kind of intelligence that is necessary for an individual to be well adjusted to its society is not easy to predict without it being situated there. Not only are there emergent societal dynamics that only occur in that context but also the society facilitates adaptive behaviors in the individual that are not possible on its own. In other words not only is society constructed by society (at least partially) but also the individual’s intelligence is so built. The authors in this section of the book are all involved in seeking to understand societies of agents alongside the individual’s social intelligence.

In chapter 10 Juliette Rouchier uses observations of human social intelligence to suggest how we might progress towards implementing a meaningful social intelligence in agents. She criticizes both the complex designed agent approach and the Artificial Life approach as failing to produce a social life that is close to that of humans, in terms of creativity or exchange of abstractions. She argues that agents will require a flexibility in communicative ability that allows to build new ways of communicating, even with unknown entities and are able to transfer a protocol from one social field to another. A consequence of this is that fixed ontologies and communication protocols will be inadequate for this task.

Hidekazu Kubota and Toyoaki Nishida (chapter 11) describe an implemented system where a number of "artificial egos" discursively interact to create community knowledge. This is a highly innovative system where the artificial egos can converse to form narratives which are relayed back to their human counterparts. The associative memory of the egos is radically different from those of traditional agents, because the idea is that the egos concentrate on the relevance of contributions rather than reasoning about the content. This structure facilitates the emergence of community knowledge. Whether or not this style of approach will turn out to be sufficient for the support of useful community knowledge, this is a completely new and bold style which will doubtlessly be highly influential on future efforts in this direction.

In chapter 12 David Pynadath and Milind Tambe report their experience in using a system of electronic assistants, in particular focusing on teams of agents operating in a real-world human organization. Their experience lead them to abandon a decision tree approach and instead adopt a more adaptive model that reasons about the uncertainty, costs, and constraints of decisions. They call this approach *adjustable autonomy* because the agents take into account the potential bad consequences of their action when deciding to take independent action, much as an employee might check critical decisions with her boss. The resulting system now assists their research group in rescheduling meetings, choosing presenters, tracking people's locations, and ordering meals.

Edmund Chattoe is a sociologist who uses agent-based computational simulation as a tool. In chapter 13 he argues that rather than basing the design of our agent systems upon a priori design principles (e.g. from philosophy) we should put considerable effort into collecting information on human society. He argues that one factor hindering realization of the potential of MAS (multi-agent systems) for social understanding is the neglect of systematic data use and appropriate data collection techniques. He illustrates this with the example of innovation diffusion and concludes by pointing out the advantages of MAS as a tool for understanding social processes.

The following 20 chapters can be thematically grouped into five sections which describe how Socially Intelligent Agents are being implemented and used in a wide range of practical applications. This part shows how Socially Intelligent Agents can contribute to areas where social interactions with humans are a necessary (if not essential) element in the commercial success and acceptance of an agent system. The chapters describe SIA systems that are used for a variety of different purposes, namely as therapeutic systems (section 2.4), as physical instantiations of social agents, namely social robots (section 2.5), as systems applied in education and training (section 2.6), as artifacts used in games and entertainment (section 2.7), and for applications used in e-commerce (section 2.8).

2.4 Interactive Therapeutic Agent Systems

Interactive computer systems are increasingly used in therapeutic contexts. Many therapy methods are very time- and labor-extensive. Computer software can provide tools that allow children and adults likewise to learn at their own pace, in this way taking some load off therapists and parents, in particular with regard to repetitive teaching sessions. Computer technology is generally very 'patient' and can easily repeat the same tasks and situations over and over again, while interaction and learning histories can be monitored and tracked. At the same time, interaction with computer technology can provide users with rewarding and often very enjoyable experiences. The use of Socially Intelligent Agents (robotic or software) in autism therapy is a quite recent development. People with autism generally have great difficulty in social interaction and communication with other people. This involves impairments in areas such as recognizing and interpreting the emotional meaning of facial expressions, difficulties in turn-taking and imitation, as well as problems in establishing and maintaining contact with other people. However, many people with autism feel very comfortable with computer technology which provides a, in comparison to interactions with people, relatively safe and predictable environment that puts the person in control. Three chapters in this section address the use of interactive agents in autism therapy from different viewpoints. The last chapter discusses the application area of providing counseling support where embodied virtual agents are part of a 'therapy session'.

Chapter 14 reports on results emerging from the project Aurora (Autonomous robotic platform as a remedial tool for children with autism). It is a highly interdisciplinary project involving computer scientists, roboticists and psychologists. Aurora is strongly therapeutically oriented and investigates systematically how to engage children with autism in interactions with a social robot. A central issue in the project is the evaluation of the interactions that occur during the trials. Such data is necessary for moving towards the ultimate goal of demonstrating a contribution to autism therapy. This chapter introduces two different techniques that assess the interactive and communicative competencies of children with autism. A quantitative technique based on micro-behaviors allows to compare differences in children's behavior when interacting with the robot as opposed to other objects. Secondly, it is shown how a qualitative technique (Conversation Analysis) can point out communicative competencies of children with autism during trials with the mobile robot.

In chapter 15 François Michaud and Catherine Théberge-Turmel describe different designs of autonomous robots that show a variety of modalities in how they can interact with people. This comprises movements as well as vocal messages, music, color and visual cues, and others. The authors goal is to engineer robots that can most successfully engage different children with autism. Given the large individual differences among people diagnosed along the autistic spectrum, one can safely predict that one and the same robot might not work with all children, but that robots need to be individually tailored towards the needs and strengths of each child. The authors' work demonstrates research along this direction to explore the design space of autonomous robots in autism therapy. The chapter describes playful interactions of autistic children and adults with different robots that vary significantly in their appearance and behavior, ranging from spherical robotic 'balls' to robots with arms and tails that can play rewarding games.

Chapter 16 discusses how an interactive computer system can be used in emotion recognition therapy for children with autism. Katharine Blocher and Rosalind W. Picard developed and tested a system called *Affective Social Quest* (ASQ). The system includes computer software as well as toy-like 'agents', i.e. stuffed dolls that serve as haptic interfaces through which the child interacts with the computer. This approach therefore nicely bridges the gap between the world of software and the embodied world of physical objects[4]. Practitioners can configure ASQ for individual children, an important requirement for the usage of computer technology in therapy. Evaluations tested how well children with autism could match emotional expressions shown on the computer screen with emotions represented by the dolls. Results of the evaluations are encouraging. However, and as it is the case for all three chapters in this book on autism therapy, the authors suggest that long-term studies are necessary in order to provide more conclusive results with regard to how interactive systems can be used in autism therapy.

In chapter 17 Stacy C. Marsella describes how socially intelligent animated virtual agents are used to create an 'interactive drama'. The drama called *Carmen's Bright IDEAS* has clear therapeutic goals: the particular application area is therapeutic counseling, namely assisting mothers whose children undergo cancer treatment in social problem solving skills. The interactive pedagogical drama involves two characters, the counselor Gina, and Carmen who represents the mother of a pediatric cancer patient. The user (learner) interacts with Gina and Carmen and it is hoped that these interactions provide a therapeutic effect. Important issues in this work are the creation of believable characters and a believable story. In order to influence the user, the system needs to engage the user sufficiently so that she truly empathizes with the characters. The system faces a very demanding audience, very different e.g. from virtual dramas enacted in game software, but if successful it could make an important contribution to the quality of life of people involved.

2.5 Socially Intelligent Robots

Embodied socially intelligent robots open up a wide variety of potential applications for social agent technology. Robots that express emotion and can cooperate with humans may serve, for example, as toys, service robots, mobile tour guides, and other advice givers. But in addition to offering practical applications for social agent technology, social robots also constitute powerful tools to investigate cognitive mechanisms underlying social intelligence. The first three chapters of this section propose robotic platforms that embed some of the cognitive mechanisms required to develop social intelligence and to achieve socially competent interactions with humans, while the fourth one is primarily concerned with understanding human response to "perceived" social intelligence in order to gain insight for the design of the socially adept artifacts of the future.

In chapter 18, Cynthia Breazeal discusses her approach to the design of sociable machines as “a blend of art, science, and engineering”, and outlines some of the lessons learned while building the sociable ‘infant’ robot Kismet. With a strong developmental approach that draws inspiration from findings in the psychology literature, combined with the idea of giving the robot an appearance that humans find attractive and believable enough to engage in infant-caregiver interactions with it, Breazeal develops four principles that guided the design of Kismet—regulation of interactions, establishment of appropriate social expectations, readable social cues, and interpretation of human social cues. Those principles provide the rationale that explains the role of the different elements engineered in Kismet’s architecture, in particular of its ‘social machinery’ and of the resulting behavior.

Chapter 19, by Hideki Kozima, presents Infanoid—an infant-like robot designed to investigate the mechanisms underlying social intelligence. Also within a developmental perspective, Kozima proposes an ‘ontogenetic model’ of social intelligence to be implemented in Infanoid so that the robot achieves communicative behavior through interaction with its social environment, in particular with its caregivers. The model has three stages: (1) the acquisition of intentionality, in order to allow the robot to make use of certain methods to attain goals; (2) identification with others, which would allow it to experience others’ behavior in an indirect way; and (3) social communication, by which the robot would understand others’ behavior by ascribing intentions to it. In this chapter, Kozima outlines some of the capabilities that Infanoid will have to incorporate in order to acquire social intelligence through those three stages.

In chapter 20, Aude Billard discusses how the Piagetian ideas about the role of ‘play, dreams, and imitation’ in the development of children’s understanding of their social world are relevant to Socially Intelligent Agents research. Billard discusses these notions in the context of the Robota dolls, a family of small humanoid robots that can interact with humans in various ways, such as imitating gestures to learn a simple language, simple melodies, and dance steps. Conceived in the spirit of creating a robot with adaptable behavior and with a flexible design for a cute body, the Robota dolls are not only a showcase of artificial intelligence techniques, but also a (now commercial) toy and an educational tool. Billard is now exploring the potential benefits that these dolls can offer to children with diverse cognitive and physical impairments, through various collaborations with educators and clinicians.

Chapter 21, by Mark Scheeff, John Pinto, Kris Rahardja, Scott Snibbe, and Robert Tow, describes research on Sparky, a robot designed with the twofold purpose to be socially competent in its interactions with humans, and to explore human response to such ‘perceived’ social intelligence, in order to use the feedback gained to design artifacts which are more socially competent in the future. Sparky is not autonomous but teleoperated, since the current state of the art in mobile and social robotics does not permit to achieve complex and rich enough interactions. In addition to facial expression, Sparky makes extensive use of its body (e.g., posture, movement, eye tracking, mimicry of people's motions) to express emotion and to interact with humans. The authors report and discuss very interesting observations of people interacting with the robot, as well as the feedback provided in interviews with some of the participants in the experiments and with the operators of Sparky.

2.6 Interactive Education and Training

Virtual training environments can provide (compared with field studies) very cost-efficient training scenarios that can be experimentally manipulated and closely monitor a human's learning process. Clearly, interactive virtual training environments are potentially much more 'engaging' in contrast to non-interactive training where relevant information is provided passively to the user, e.g. in video presentations. The range of potential application areas is vast, but most promising are scenarios that would otherwise (in real life) be highly dangerous, cost-intensive, or demanding on equipment.

Similarly, Socially Intelligent Agents in children's (or adult's) education can provide enjoyable and even entertaining learning environments, where children learn constructively and cooperatively. Such learning environments cannot replace 'real life' practical experience, but they can provide the means to creatively and safely explore information and problem spaces as well as fantasy worlds. Using such environments in education also provides useful computer skills that the children acquire 'by doing'. Education in such systems can range from learning particular tasks (such as learning interactively about mathematics or English grammar), encouraging creativity and imagination (e.g. through the construction of story environments by children for children), to making a contribution to personal and social education, such as getting to know different cultures and learning social skills in communication, cooperation and collaboration with other children that might not be encountered easily in real life (e.g. children in other countries).

In chapter 22 Jonathan Gratch describes 'socially situated planning' for deliberate planning agents that inhabit virtual training environments. For training simulators, in order to be believable, not only the physical dynamics, but also the social dynamics and the social behavior of the agents must be designed carefully. For learning effects to occur, such training scenarios need to be 'realistic' and believable enough to engage the user, i.e. to let the user suspend the disbelief that this is not 'just a simulation' where actions do not matter. In the proposed architecture, social reasoning is realized as a meta-level on top of a general purpose planning layer. The system's capabilities are illustrated with interactions between two synthetic characters, Jack and Steve, who have conflicting goals. Changing variables in the system leads to different types of interactions, rude as opposed to cooperative interaction. While subtleties of social behavior cannot be modeled, experience in real-world military simulation applications suggests that some social interactions can be modeled adequately.

Chapter 23 discusses the design of empathic ambience in the context of computer-based learning environments for children. A key factor in human social understanding and communication is empathy which helps people to understand each other's perspectives, and to develop their own perspectives. Bridget Cooper and Paul Brna argue that the ambience in learning environments depends on the quality of communication and interaction. This ambience can be supported by empathic design which takes into account interactions, emotions, communication and social relationships. A 'pedagogical claims analysis' (a participatory design) methodology is used in the evaluation of the design process, involving both teachers and pupils. The chapter discusses the design and support of empathy and reports on work that studies the role of empathy in teacher/pupil relationships. Results in classrooms suggest that the approach taken created a positive model of how teachers and children can work together with computers in the classroom setting.

In chapter 24 Isabel Machado and Ana Paiva describe some design decisions taken in the construction of a virtual story-creation environment called *Teatrix*. In Teatrix children can collaboratively create and reflect upon virtual stories. Story-telling is not only an enjoyable activity for children (and adults) but also an important element in a child's cognitive and social development. Each character in the virtual game has a certain role and a certain function in the story. Children can control the characters which can also act autonomously. Children can communicate through their characters by letting them interact or 'talk' to each other. Tests with children showed the need for a higher level of understanding of the characters' behavior. This led to the development of a meta-level control tool called 'hot seating'. Here, children take the character's viewpoint and have to justify its behavior which can give children a chance to reflect on and better understand the character's actions.

Chapter 25 describes work done by an intergenerational design team where children are design partners in the construction of new story-telling technology for children. Such technology includes the emotional robotic storyteller *PETS* and the construction kit *Storykit* that allows children to build interactive physical story environments. Jaime Montemayor, Allison Druin and Jim Hendler use the design methodology of 'cooperative inquiry' where children are included as design partners. PETS is a robotic story-telling system that elementary school age children can use to build their own robotic animal pet by connecting body parts. A particular software (*My PETS*) can be used to write and tell stories and to create ‘emotions’ that the robot can act out. Using Storykit children can create their own StoryRooms that provide story-telling experience. Tests of PETS and StoryKit were promising and let to a list of design guidelines that for building attractive and interactive story environments for children.

2.7 Socially Intelligent Agents in Games and Entertainment

This section concerns important mainstream applications of the technology of socially intelligent agents, in educational games, in interactive drama, and in interactive art. In educational games, agents must exhibit enough social sophistication so as to be able to flexibly manage students’ emotional states and learning engagement. In a drama of purely autonomous agents, each agent would need to be equipped with sufficient intelligence to react reasonably to the range of situations that can occur; those that can be generated by the total system. This intelligence presumably is represented in the form of social knowledge, abilities for perceiving and understanding other’s behaviors, the ability to identify and characterize problems, and the ability to generate and execute plans for solving these goals. In order to make this enormous problem tractable, we can limit the range of possibilities to certain classes of behaviors, social interactions and goals. Although the agents stay within a given class of behaviors, an observing human will perceive an extended range of intentions. When we then try to involve a human in an agent drama, we have to provide for agents perceiving the actions of the human. More importantly, the human will not be able to stay within a prespecified class of behaviors. Thus, agents will need to respond to a wider range of actions and situations. This presents a major challenge for agent designers. Further, we will usually want more of the ensuing action than the human just spending time in the virtual social world. We want to arrange for the human to take part in a *drama* with certain dramatic goals which express the author’s intent. Thus, in interactive drama we hit core issues of the development of characters which can dynamically respond to novel situations in ways which are not only socially appropriate but which further dramaturgic goals. In interactive art, we descend into the self of the human interactor.

In chapter 26, Cristina Conati and Maria Klawe explain how the flexibility and social appropriateness achievable with socially intelligent agents can effectively support the learning process of students. They describe their system for multiplayer multiactivity educational games. The main issues concern how socially intelligent agents can model the players’ cognitive and metacognitive skills, i.e. including their management of their own cognitive activity, as well as motivational states and engagement in a collaborative interaction.

In chapter 27, Michael Mateas and Andrew Stern describe their approach to building an interactive drama system in which a human user participates in a dramatic story and thereby experiences it from a first person perspective. The main problem is to design agents with less than human abilities but which can nevertheless play believable roles in a range of situations. Their approach is to provide a drama manager agent which keeps the overall action on course, and also thereby reduces the demands on characters who therefore need only use local plans applicable in the vicinity of the story line.

Michael Young discusses another approach to interactive drama in chapter 28. The narrative structure of the games is generated dynamically, and its main principle is to manage a cooperative contract with the user. This consists of dramatical expectations built upon social commitments. The system creates, modifies and maintains a narrative plan using dramatical principles, and the unfolding of action is designed to provide an interesting narrative experience for the user.

In chapter 29 Nell Tenhaaf manages to bring together the treatments of self for interactive agents produced by artists for interactive art and those produced by computer scientists for intelligent agent applications. Her discussion illuminates the depth of this subject and points us to its sophisticated literature. She also describes in detail one particular interactive work entitled ‘Talk Nice’ made by fellow artist Elizabeth Van Der Zaag. Using video and a speech recognition system, this implements a bar ‘pick up’ social situation where the user has to talk nice to succeed.

2.8 Social Agents in E-Commerce

It is not surprising to find a section of this book dealing with commerce, since the exchange of value is one of the principle social mechanisms humans use. In the last century economics tried to strip exchange of its social aspects by the use of strong normative assumptions. Their models insisted (in practice) of very limited and selfish goals for its agents, they limited communication to the barest minimum (usually to price alone) and they almost totally ignored any process preferring to concentrate on equilibrium states instead. Now that it is becoming increasingly clear that this approach has failed, there is a renewed interest in using MAS to model these processes – putting some of the critical aspects that were jettisoned back in. At the same time the exchange of value is being increasingly conducted using computational media. The effect of this is to somewhat disembody the exchange process which makes it possible for software agents to participate as near equals with humans. The confluence of using societies of agents to model the complexities of social exchange and the challenge of using them to perform that exchange reinforces the importance social agents will have with respect to commerce in the next century.

In chapter 30, Peyman Faratin considers the relationship between knowledge, computation and the quality of solution for an agent involved in negotiation. Starting from a fairly classical game-theory model he relaxes the assumptions in order to approach the situation real computational agents will find themselves in. His results indicate that the type of cognitive model that the agents have in a negotiation substantially effects the outcome and he concludes that learning is an important skill for an agent involve in a realistic negotiation.

Scott Moss (chapter 31) uses agent-based simulations to try to understand social systems. This paper is an interim report on an attempt to understand negotiation between humans by investigating negotiation between agents. He grounds his model with a real example of negotiation: the multi-party negotiation between the various parties interested in the Meuse river. In this model agents negotiation over a multi-dimensional space of possibilities where each agent will not only have different goals but also attach different importance to different goals. His agents learn who to negotiate with based upon observations of the other agents with respect to properties such as: trustworthiness, reliability and similarity. His result is that although two agents succeed three or more fail. This indicates that coalitions of agents might be critical to the success of any multi-party negotiation (as well as the difficulty of the task).

In chapter 32 Juan A. Rodríguez-Aguilar and Carles Sierra start from a macro perspective to try and design "organization centered" MAS. Like Scott Moss they do not start from traditional a priori models, but take a real human example (in this case a fish market) as their guide. From this they abstract what they see as the principle institutional components and show how this can lead to an effective open and agent-mediated institution. They claim that claim that such a computational model is general enough to found the development of other agent institutions.

The last chapter of the book (33) by Helen McBreen is an empirical study of the reaction of people to virtual sales assistants. These assistants are 3D embodied conversational agents that interact with a customer. She evaluated customers' reactions in three interactive VRML e-commerce environments: a cinema box office, a travel agency and a bank. She found that the customers carried over their expectations in terms of dress from the real world and that they found it hard to trust the banking agent.

3. Common Themes

As mentioned above, many themes that are addressed in the 33 chapters apply across different chapters. A few selected themes are listed in Figure 1.3. This 'mental map' might help readers with specific interests in navigating the book.

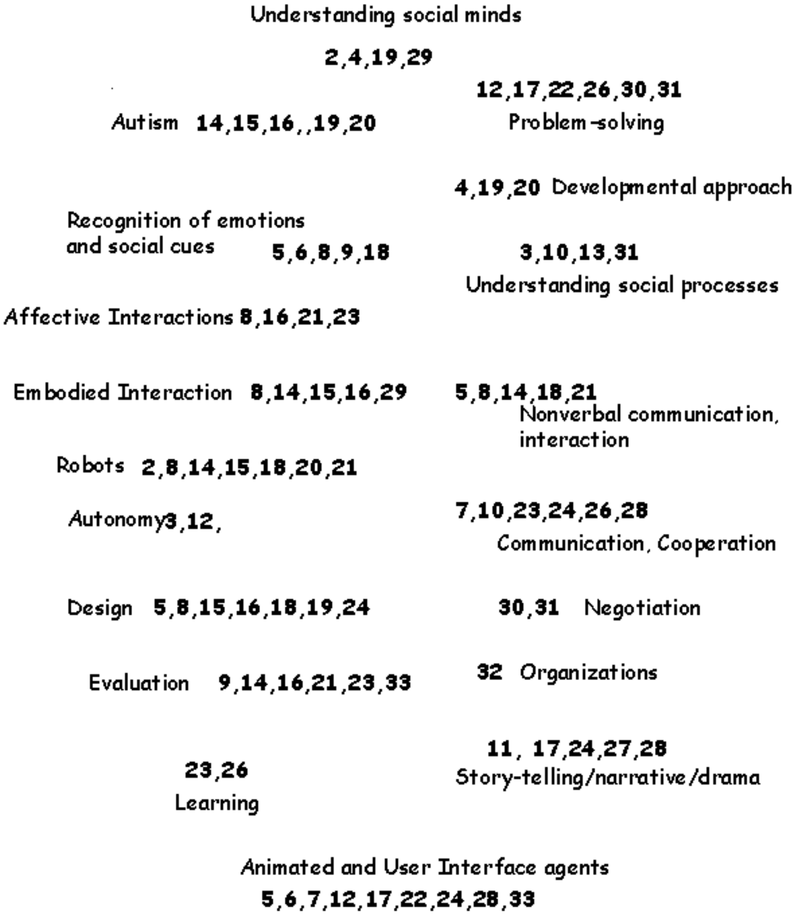

Understanding social minds
2,4,19,29
12,17,22,26,30,31
Problem-solving
Autism 14,15,16,,19,20
4,19,20 Developmental approach
Recognition of emotions
and social cues 5,6,8,9,18
3,10,13,31
Understanding social processes
Affective Interactions 8,16,21,23
Embodied Interaction 8,14,15,16,29
5,8,14,18,21
Nonverbal communication,
interaction
Robots 2,8,14,15,18,20,21
Autonomy3,12,
7,10,23,24,26,28
Communication, Cooperation
Design 5,8,15,16,18,19,24
30,31 Negotiation
Evaluation 9,14,16,21,23,33
32 Organizations
11, 17,24,27,28
Story-telling/narrative/drama
23,26
Learning
Animated and User Interface agents
5,6,7,12,17,22,24,28,33

Figure 1.3. Selected themes that apply across section boundaries.

Acknowledgments

We would like to thank Gerhard Weiss, the series editor of *Multiagent Systems, Artificial Societies, and Simulated Organizations* for the exciting opportunity to publish this book. We also thank Melissa Fearon from Kluwer Academic Publishers for continuous advice during the editing and publication process. For their support of the AAAI Fall symposium Socially Intelligent Agents – The Human in the Loop, from which this book emerged, we like to thank AAAI (the American Association for Artificial Intelligence). Kerstin Dautenhahn, the chair of the AAAI Fall Symposium, warmly thanks the co-organisers: Elisabeth André (DFKI GmbH, Germany), Ruth Aylett (Univ. Salford, UK), Cynthia Breazeal (MIT Media Lab, USA), Cristiano Castelfranchi (Italian National Research Council, Italy), Justine Cassell (MIT Media Lab, USA), Francois Michaud (Univ. de Sherbrooke, Canada), and Fiorella de Rosis (Univ. of Bari,

Italy). Maria Miceli (Italian National Research Council, Italy) and Paola Rizzo (Univ. of Rome "La Sapienza", Italy) kindly acted as additional reviewers for the 2000 AAAI Fall Symposium.

Notes

1. Examples of collections of articles on SIA research in book and special journal issues are: K.Dautenhahn, C. Numaoka (guest editors): Socially Intelligent Agents, Special Issues of *Applied Artificial Intelligence*, Vol. 12 (7-8), 1998, and Vol. 13(3), 1999, K.Dautenhahn (2000): *Human Cognition and Social Agent Technology*, John Benjamins Publishing Company, B. Edmonds and K. Dautenhahn (guest editors): Social Intelligence, special issue of *Computational and Mathematical Organisation Theory*, Vol. 5(3), 1999, K. Dautenhahn (guest editor): Simulation Models of Social Agents, special issue of *Adaptive Behavior*, Vol. 7(3-4), 1999, Bruce Edmonds and Kerstin Dautenhahn (guest editors): Starting from Society - the application of social analogies to computational systems, special issue of *The Journal of Artificial Societies and Social Simulation (JASSS)*, 2001. Kerstin Dautenhahn (guest editor): Socially Intelligent Agents – The Human in the Loop, special issue of *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, Vol. 31(5), 2001; Lola Cañamero and Paolo Petta (guest editors), Grounding emotions in adaptive systems, special issue of *Cybernetics and Systems*, Vol. 32(5) and Vol. 32(6), 2001.

2. see events listed on the SIA Webpage: http://homepages.feis.herts.ac.uk/ comqkd/aaai-social.html

3. Guest Editor: Kerstin Dautenhahn. Table of Contents: *Guest Editorial: Socially Intelligent Agents - The Human in the Loop* by Kerstin Dautenhahn; *Understanding Socially Intelligent Agents – A Multi-Layered Phenomenon* by Per Persson, Jarmo Laaksolahti, Peter Lönnqvist; *The child behind the character* by Ana Paiva, Isabel Machado, Rui Prada, *Agents supported adaptive group awareness: Smart distance and WWWare* by Yiming Ye, Stephen Boies, Paul Huang, John K. Tsotsos; *Socially intelligent reasoning for autonomous agents* by Lisa Hogg and N. Jennings; *Evaluating humanoid synthetic agents in e-retail applications* by Helen McBreen, Mervyn Jack, *The Human in the Loop of a Delegated Agent: The Theory of Adjustable Social Autonomy* by Rino Falcone and Cristiano Castelfranchi; *Learning and Interacting in Human-Robot Domains* by Monica N. Nicolescu and Maja J. Matari¢; *Learning and communication via imitation: an autonomous robot perspective* by P. Andry, P. Gaussier, S. Moga, J. P. Banquet, J. Nadel; *Active vision for sociable robots* by Cynthia Breazeal, Aaron Edsinger, Paul Fitzpatrick, Brian Scassellati; *I Show You How I Like You: Can You Read it in My Face?* by Lola D. Cañamero, Jakob Fredslund; *Diminishing returns of engineering effort in telerobotic systems* by Myra Wilson, Mark Neal and *Let's Talk! Socially Intelligent Agents for Language Conversation Training* by Helmut Prendinger, Mitsuru Ishizuka.

4. Compare [8] for teaching the recognition and understanding of emotions and mental states.

References

[1] Jeffrey M. Bradshaw, editor. *Software Agents*. AAAI Press/The MIT Press, 1997.

[2] Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, editors. *Embodied conversational agents*. MIT Press, 2000.

[3] K. Dautenhahn, editor. *Human Cognition and Social Agent Technology*. John Benjamins Publishing Company, 2000.

[4] K. Dautenhahn and C. L. Nehaniv, editors. *Imitation in Animals and Artifacts*. MIT Press (in press), 2002.

[5] Kerstin Dautenhahn. The art of designing socially intelligent agents: science, fiction and the human in the loop. *Applied Artificial Intelligence Journal, Special Issue on Socially Intelligent Agents*, 12(7-8):573–617, 1998.

[6] Mark D'Inverno and Michael Luck, editors. *Understanding Agent Systems*. The MIT Press, 2001.

[7] Allison Druin and James Hendler, editors. *Robots for Kids – Exploring new technologies for learning*. Morgan Kaufmann Publishers, 2000.

[8] Patricia Howlin, Simon Baron-Cohen, and Julie Hadwin. *Teaching Children with Autism to Mind-Read.* John Wiley and Sons, 1999.

[9] Michael N. Huhns and Munindar P. Singh, editors. *Readings in Agents*. Morgan Kaufmann Publishers, Inc., 1998.

[10] Ana Paiva, editor. *Affective Interactions*. Springer-Verlag, 2000.

[11] Phoebe Sengers and Michael Mateas, editors. *Narrative Intelligence*. John Benjamins Publishing Company (to appear), 2002.

[12] Robert Trappl and Paolo Petta, editors. *Creating personalities for synthetic actors*. Springer Verlag, 1997.

Chapter 2

UNDERSTANDING SOCIAL INTELLIGENCE

Per Persson[1], Jarmo Laaksolahti[1] and Peter Lönnqvist[2]
[1] *Swedish Institute of Computer Science, Kista, Sweden,* [2] *Department of Computer and Systems Sciences, Stockholm University and Royal Institute of Technology*

Abstract Believable social interaction is not only about agents that look right but also do the right thing. To achieve this we must consider the everyday knowledge and expectations by which users make sense of real, fictive or artificial social beings. This folk-theoretical understanding of other social beings involves several, rather independent, levels such as expectations on behaviour, expectations on primitive psychology, models of folk-psychology, understanding of traits, social roles and empathy. Implications for Socially Intelligent Agents (SIA) research are discussed.

1. Introduction

Agent technology refers to a set of software approaches that are shifting users' view of information technology from tools to actors. Tools react only when interacted with, while agents act autonomously and proactively, sometimes outside the user's awareness. With an increasing number of autonomous agents and robots making their way into aspects of our everyday life, users are encouraged to understand them in terms of human behaviour and intentionality. Reeves and Nass [5] have shown that people relate to computers - as well as other types of media - as if they were 'real', e.g., by being polite to computers. However, some systems seem to succeed better than others in encouraging such anthropomorphic attributions, creating a more coherent and transparent experience [20]. What are the reasons for this? What encourages users to understand a system in terms of human intentionality, emotion and cognition? What shapes users' experiences of this kind? Software agent research often focuses on the graphical representation of agents. Synchronisation of lip movements and speech, gestures and torso movements as well as the quality of the graphical output itself are questions that have been investigated [6] [14]. In contrast, the authors of this chapter propose a multi-facetted view of how users employ an intentional stance in understanding socially intelligent agents.

In order to understand how and why users attribute agents with intelligence in general and social intelligence in particular, to we turn to a *constructivist explanation model*. The ontological claims underlying this approach focus mainly on the active role of the human mind in constructing a meaningful reality [25]. 'Social intelligence' is not some transcendental faculty, but an understanding arising in the interaction between a set of cues and an active and cognitively creative observer. Thanks to the constructively active user, the cues needed to prompt anthropomorphic attributions can be quite simple on the surface [1] [5, p. 7] [27, p. 173].

Since science knows little about how 'real' intelligence, intentionality or agency work - or even if there are such things outside of human experience - we cannot create intelligence independently of an observer/user. In order to achieve appearance of intelligence it is crucial to design SIA systems with careful consideration to how such systems will be received, understood and interpreted by users. The function of SIA technology becomes the centre of attention, whether this is learning [30], therapy [19], game/play experiences [22] [15], the SIMS or the spectacular appearance of a Sony Aibo robotic dog. According to a constructivist approach to SIA, there is little use in creating artificial intelligence unless it is meaningful consistent [20] and coherent to a given user.

An opposing view of social intelligence research takes an *objectivist standpoint*. According to this view - rooted in strong AI - social intelligence is something that can be modelled and instantiated in any type of hardware, software or wetware, but transcendentally exists outside any such instantiation. The aim is to create SIA that are socially intelligent in the same sense as humans are and thus the models created are based on theories of how actual human social intelligence manifests itself.

Depending on the view taken the purpose of SIA research differs. While constructivists aim to study how users understand, frame and interpret intelligent systems in different situations, and use this knowledge to improve or enhance the interaction, objectivists aim to study emergent behaviour of systems and find better models and hypotheses about how human intelligence works.

The purpose of this chapter is to develop a conceptual framework, describing how understandings/impressions of social intelligence arise in users. Once this is in place, we will be able to develop a method for investigating and developing socially intelligent agents.

2. Folk-Theories: 'Naive' Theories about Intelligence

There is reason to believe that people employ the same or similar psychological and social strategies when making sense of artificially produced intelligent behaviour as with real world intelligence (e.g., humans and animals). There might be some minor variations in reception dependent on media (computer, theatre, film or in everyday situations), or if the intelligence is thought to be fictive/simulated or real/documentary - but the major bulk of employed psychosocial skills will overlap (in the case of cinema characters, see [25]). We will call such skills *folk-theories*, since they are knowledge and hypotheses about the world, albeit of a 'naive' and common-sense nature. People and cultures employ such naive theories in many areas of everyday life, e.g., physics, nature, psychology, energy, morality, causality, time and space [12]; [9]. For our purposes, we will deal only with folk-theories about intelligent behaviour, interpersonal situations, and social reality.

Although people have idiosyncratic expectations about intelligent behaviour, for instance specific knowledge about the personality and habits of a close friend, folk-theories constitute the collectively shared knowledge in a social, cultural or universal group of people. Folk-theories constitute users' expectations about intelligent behaviour. In order for the system to appear intelligent, it must meet those expectations, at least on some level.

Elsewhere we have described these folk-theories in detail and given examples of SIA systems that seek to accommodate these [26]. Here space allows only a brief overview.

2.1 Examples of Folk-Theories

If intelligence is embodied in some form, then people have expectations about visual appearance and physical behaviour. People have visual expectations of bodies' configuration, arrangement and movement patterns, both in humans and other forms of intelligent life [10]. People expect gestures and non-verbal behaviour to be synchronized and appropriate to the situation in which they occur [24] [6]. Behaviour related to gazing and personal space is also expected to take place according to certain norms and conventions [7].

Surface behaviour of this kind, however, is never understood on its own. Users will always try to make sense of such behaviour in more abstract terms. *Primitive psychology* is a folk-theory about how basic needs such as hunger, thirst, sexual drives, and pain work, and the different ways in which they are related (e.g., hunger or thirst will disappear if satisfied, and that satisfaction will fade over time until hunger or thirst reoccur). *Folk-psychology* constitutes a common sense model about how people understand the interrelationships between different sorts of mental states in other people (and in themselves), and how these can be employed as common-sense explanations for external behaviour and action. Research on an 'everyday theory of mind', for instance, studies how people relate perceptions, thinking, beliefs, feelings, desires, intentions and sensations, and reason about these [2] [18] [29] [17] [16]. The ways in which people attribute and reason about emotions of other people have been studied within appraisal theory [13] [28] [31] - for overview, see [4].

At yet a higher level, people understand intelligent behaviour in terms of *personality*, which refers to dimensions of a person that are assumed to be more stable and enduring than folk-psychological mental states. People may, for instance use a common-sense theory about *traits* to explain the behaviour of other people [23] (Per's tendency to be late is often explained by Jarmo and Peter by referring to 'his carelessness'). People also have sophisticated folk-theories about social roles and expectations about the behaviours of these roles in specific situations, for instance family roles (father, mother, daughter), occupancy roles (fireman, doctor, waiter), social stereotypes, gender stereotypes, ethnic stereotypes or even archetypes of fictions and narratives (the imbecile, the hypochondriac, Santa Clause). Social roles are studied within social psychology, sociology, anthropology, ethnology and communication studies e.g., [32, p. 91] [21, p. 39].

In addition to these folk-theories, people also expect intelligent agents not only to be responsive to input, but to proactively take action on the basis of the agent's assumed goals, desires and emotions - cf. Dennett's, [8] distinction between *mechanical* and *intentional stance*. To a certain extent we also expect intelligent agents to be able to *learn* new things in light of old knowledge, or to apply old knowledge to new contexts. This, in fact, seems to be one of the central features of human intelligence.

Finally, people expect intelligent creatures to pay special attention to other intelligent creatures in the environment, and be able to relate to the point of view of those individuals. Defined broadly, people expect intelligent creatures to have emphatic capabilities (cf. [4]). This may include perceptual processes (being able to follow the user's gaze; cf., [11], cognitive processes (inferring the goals and emotions of the user) as well as 'true' emotional empathy (not only attributing a mental state to a person, but also *sharing* that emotion or belief, or some congruent one).

2.2 Features of Folk-Theories

Folk-theories about social intelligence are not idiosyncratic bits and pieces of common sense wisdom, but constitute coherent cognitive networks of interrelated entities, shared by a large number of people. Folk-theories are structures that organize our understanding and interaction with other intelligent creature. If a given behaviour can be understood in terms of folk-theoretical expectations, then it is experienced as 'meaningful'. If some aspect of the situation falls outside the interrelationships of the folk-theories, then the behaviour is judged to be 'incomprehensible', 'strange', 'crazy' or 'different' in some form. This often happens in, for instance, inter-cultural clashes. Albeit such misunderstandings are due to social and cultural variations of folk-theories, most folk-theories probably possess some form of universal core shared by all cultures [25, p. 226].

From an evolutionary point of view, folk-theories about intelligence are quite useful to an organism, since their structured nature enables reasoning and predictions about future behaviour of other organisms (see e.g. [2]). Such predictions are naive and unreliable, but surely provide better hypotheses than random guesses, and thus carry an evolutionary value.

Folk-theories are not static but change and transform through history. The popularised versions of psychoanalysis, for instance, perhaps today constitute folk-theoretical frameworks that quite a few people make use of when trying to understand the everyday behaviours of others.

Folk-theories are acquired by individuals on the basis of first-person deduction from encounters with other people, but perhaps more importantly from hearsay, mass-media and oral, literary and image-based narratives [3] [9].

In summary, folk-theories about social intelligence enable and constrain the everyday social world of humans.

3. Implications for AI Research

If users actively attribute intelligence on the basis of their folk-theories about intelligence, how will this affect they way in which SIA research is conducted? First, in order to design apparently intelligent systems, SIA researchers need not study scientific theories about the mechanisms of 'real' intelligence, agency and intentionality, but rather how users think social intelligence works. This implies taking more inspiration from the fields of anthropology, ethnology, social psychology, cultural studies and communication studies. These disciplines describe the ways in which people, cultures and humanity as a whole use folk-theoretical assumptions to construct their experience of reality. Of course, sometimes objectivist and constructivist views can and need to be successfully merged, e.g., when studies of folk-theories are lacking. In these cases, SIA researchers may get inspiration from 'objectivist' theories in so far as these often are based on folk-theories [12, p. 337ff]. In general we believe both approaches have their merits giving them reason to peacefully co-exist.

Second, once the structure of folk-theories has been described, SIA research does not have to model levels that fall outside of this structure. For instance, albeit the activity of neurons is for sure an enabler for intelligence in humans, this level of description does not belong to people's everyday understanding of other intelligent creatures (except in quite specific circumstances). Hence, from the user's perspective simulating the neuron level of intelligence is simply not relevant. In the same spirit, researchers in sociology may explain people's intelligent behaviour in terms of economical, social and ideological structures, but since these theories are not (yet) folk-theories in our sense of the term, they may not contribute very much to user-centred SIA research. Again, since the focus lies on folk-theories, some scholarly and scientific theories will not be very useful. In this sense, constructivist SIA research adopts a sort of 'black-box' design approach, allowing tricks and shortcuts as long as they create a meaningful and coherent experience of social intelligence in the user.

This does not mean that the constructivist approach is only centred on surface phenomena, or that apparent intelligence is easy to accomplish. On the contrary, creating an apparently intelligent creature, which meets the user's folk-theoretical expectations and still manages to be deeply interactive, seems to involve high and yet unresolved complexity. It is precisely the interactive aspect of intelligence that makes it such a difficult task. When designing intelligent characters in cinema, for instance, the filmmakers can determine the situation in which a given behaviour occurs (and thus make it more meaningful) because of the non-interactive nature of the medium. In SIA applications, the designer must foresee an almost infinitive number of interactions from the user, all of which must generate a meaningful and understandable response form the system's part. Thus, interactivity is the real 'litmus test' for socially intelligent agent technology.

Designing SIA in the user centred way proposed here is to design social intelligence, rather than just intelligence. Making oneself appear intelligible to one's context is an inherently social task requiring one to follow the implicit and tacit folk-theories regulating the everyday social world.

References

[1] An Experimental Study of Apparent Behavior. F. Heider and M. Simmel. *American Journal of Psychology*, 57:243–259, 1944.

[2] Andrew Whiten. *Natural Theories of Mind. Evolution, Development and Simulation of Everyday Mindreading*. Basil Blackwell, Oxford, 1991.

[3] Aronson. *The Social Animal, Fifth Edition.* W. H. Freeman, San Francisco, 1988.

[4] B. L. Omdahl. *Cognitive Appraisal, Emotion, and Empathy.* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1995.

[5] B. Reeves and C. Nass. *The Media Equation.* Cambridge University Press, Cambridge, England, 1996.

[6] C. Pelachaud and N. I. Badler and M.Steedman. Generating Facial Expression for Speech. *Cognitive Science*, 20:1–46, 1996.

[7] Chris Kleinke. Gaze and Eye Contact: A Research Review. *Psychological Bulletin*, 100:78–100, 1986.

[8] D. C. Dennett. *The intentional stance.* M.I.T. Press, Cambridge, Massachusetts, 1987.

[9] Dorothy Holland and Naomi Quinn. *Cultural Models in Language and Thought*. Cambridge University Press, Cambridge, England, 1987.

[10] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201–211, 1973.

[11] George Butterworth. The Ontogeny and Phylogeny of Joint Visual Attention. In A. Whiten, editor, *Natural Theories of Mind*, pages 223–232. Basil Blackwell, Oxford, 1991.

[12] George Lakoff and Mark Johnson. *Philosophy in the Flesh. The Embodied Mind and its Challenge to Western Thought*. Basic Books, New York, 2000.

[13] I. Roseman and A. Antoniou and P. Jose. Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. *Cognition and Emotion*, 10:241–277, 1996.

[14] J. Cassell and T. Bickmore and M. Billinghurst and L. Campbell and K. Chang and H. Vilhjálmsson and H. Yan. Embodiment in Conversational Interfaces: Rea. In *ACM CHI 99 Conference Proceedings, Pittsburgh, PA*, pages 520–527, 1999.

[15] J. Laaksolahti and P. Persson and C. Palo. Evaluating Believability in an Interactive Narrative. In *Proceedings of The Second International Conference on Intelligent Agent Technology (IAT2001), October 23-26 2001, Maebashi City, Japan*. 2001.

[16] J. Perner and S. R. Leekham and H. Wimmer. Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5:125–137, 1987.

[17] J. W. Astington. *The child's discovery of the mind*. Harvard University Press, Cambridge, Massachusetts, 1993.

[18] K. Bartsch and H. M. Wellman. *Children talk about the mind*. Oxford University Press, Oxford, 1995.

[19] K. Dautenhahn. Socially Intelligent Agents and The Primate Social Brain - Towards a Science of Social Minds. In *AAAI Fall symposium, Socially Intelligent Agents - The Human in the Loop, North Falmouth, Massachusetts*, pages 35–51, 2000.

[20] Katherine Isbister and Clifford Nass. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human Computer Studies*, pages 251–267, 2000.

[21] M. Augoustinos and I. Walker. *Social cognition: an integrated introduction*. Sage, London, 1995.

[22] M. Mateas and A. Stern. Towards Integrating Plot and Character for Interactive Drama. In *AAAI Fall Symposium, Socially Intelligent Agents - The Human in the Loop, North Falmouth, MA*, pages 113–118, 2000.

[23] N. Cantor and W. Mischel. Prototypes in Person Perception. In L. Berkowitz, editor, *Advances in Experimental Psychology, volume 12*. Academic Press, New York, 1979.

[24] P. Ekman. The argument and evidence about universals in facial expressions of emotion. In *Handbook of Social Psychophysiology*. John Wiley, Chichester, New York, 1989.

[25] P. Persson. *Understanding Cinema: Constructivism and Spectator Psychology*. PhD thesis, Department of Cinema Studies, Stockholm University, 2000. (at http://www.sics.se/ perp/.

[26] P. Persson and J. Laaksolahti and P. Lonnquist. Understanding Socially Intelligent Agents - A multi-layered phenomenon, 2001. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, special issue on "Socially Intelligent Agents - The Human in the Loop",forthcoming.

[27] Paola Rizzo. Why Should Agents be Emotional for Entertaining Users? ACrtical Analysis. In Paiva, editor, *Affective Interactions. Towards a New Generation of Computer Interfaces*, pages 161–181. Springer-Verlag, Berlin, 2000.

[28] Paul Harris. Understanding Emotion. In Michael Lewis and Jaenette Haviland, editor, *Handbook of Emotions*, pages 237–246. The Guilford Press, New York, 1993.

[29] Roy D'Andrade. A folk model of the mind. In Dorothy Holland and Naomi Quinn, editor, *Cultural Models in Language and Thought*, pages 112–148. Cambridge University Press, Cambridge, England, 1987.

[30] S. Marsella. Pedagogical Soap. In *AAAI Fall Symposium, Socially Intelligent Agents - The Human in the Loop, North Falmouth, MA*, pages 107–112, 2000.

[31] S. Planalp and V. DeFrancisco and D. Rutherford. Varieties of Cues to Emotion Naturally Occurring Situations. *Cognition and Emotion*, 10:137–153, 1996.

[32] S. Taylor and J. Crocker. Schematic Bases of Social Information Processes. In T. Higgins and P. Herman and M. Zanna, editor, *Social Cognition*, pages 89–134. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1981.

Chapter 3

MODELING SOCIAL RELATIONSHIP

An Agent Architecture for Voluntary Mutual Control

Alan H. Bond
California Institute of Technology

Abstract We describe an approach to social action and social relationship among socially intelligent agents [4], based on mutual planning and mutual control of action. We describe social behaviors, and the creation and maintenance of social relationships, obtained with an implementation of a biologically inspired parallel and modular agent architecture. We define voluntary action and social situatedness, and we discuss how mutual planning and mutual control of action emerge from this architecture.

1. The Problem of Modeling Social Relationship

Since, in the future, many people will routinely work with computers for many hours each day, we would like to understand how working with computers could become more natural. Since humans are social beings, one approach is to understand what it might mean for a computer agent and a human to have a social relationship.

We will investigate this question using a biologically and psychologically inspired agent architecture that we have developed. We will discuss the more general problem of agent-agent social relationships, so that the agent architecture is used both as a model of a computer agent and as a model of a human user.

What might constitute social behavior in a social relationship? Theoretically, social behavior should include: (i) the ability to act in compliance with a set of social commitments [1], (ii) the ability to negotiate commitments with a social group (where we combine, for the purpose of the current discussion, the different levels of the immediate social group, a particular society, and humanity as a whole), (iii) the ability to enact social roles within the group, (iv) the ability to develop joint plans and to carry out coordinated action, and (v) the ability to form persistent relationships and shared memories with other individuals.

There is some systematic psychological research on the dynamics of close relationships, establishing for example their connection with attachment [5]. Although knowledge-based cognitive approaches have been used for describing discourse, there has not yet been much extension to describing relationships [6].

Presumably, a socially intelligent agent would recognize you to be a person, and assign a unique identity to you. It would remember you and develop detailed knowledge of your interaction history, what your preferences are, what your goals are, and what you know. This detailed knowledge would be reflected in your interactions and actions. It would understand and comply with prevailing social norms and beliefs. You would be able to negotiate shared commitments with the agent which would constrain present action, future planning and interpretation of past events. You would be able to develop joint plans with the agent, which would take into account your shared knowledge and commitments. You would be able to act socially, carrying out coordinated joint plans together with the agent.

We would also expect that joint action together with the agent would proceed in a flexible harmonious way with shared control. No single agent would always be in control, in fact, action would be in some sense voluntary for all participants at all times.

To develop concepts and computational mechanisms for all of these aspects of social relationship among agents is a substantial project. In this paper, we will confine ourselves to a discussion of joint planning and action as components of social behavior among agents. We will define what voluntary action might be for interacting agents, and how shared control may be organized. We will conclude that in coordinated social action, agents voluntarily maintain a regime of mutual control, and we will show how our agent architecture provides these aspects of social relationship.

2. Our Agent Architecture

In this section we describe of an agent architecture that we have designed and implemented [2] [3] and which is inspired by the primate brain. The overall behavioral desiderata were for an agent architecture for real-time control of an agent in a 3D spatial environment, where we were interested in providing from the start for joint, coordinated, social behavior of a set of interacting agents.

Data types, processing modules and connections. Our architecture is a set of processing modules which run in parallel and intercommunicate. We diagram two interacting agents in the figure. This is a totally distributed architecture with no global control or global data. Each module is specialized to process only data of certain datatypes specific to that module. Modules are connected by a

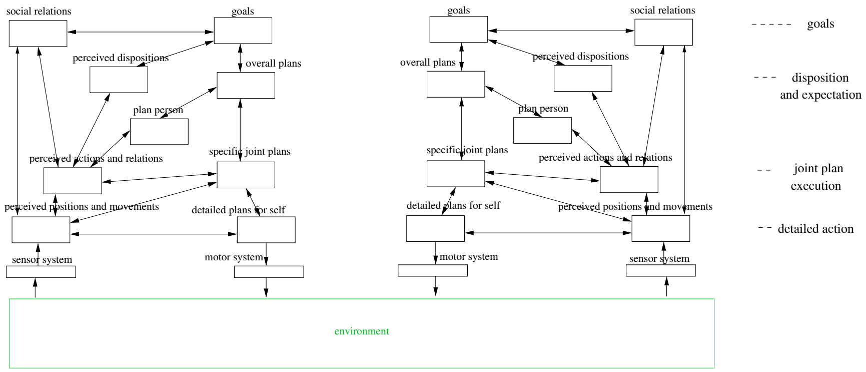


Figure 3.1. Our agent architecture

fixed set of connections and each module is only connected to a small number of other modules. A module receives data of given types from modules it is connected to, and it typically creates or computes data of other types. It may or may not also store data of these types in its local store. Processing by a module is described by a set of left-to-right rules which are executed in parallel. The results are then selected competitively depending on the data type. Typically, only the one strongest rule instance is allowed to "express itself", by sending its constructed data items to other modules and/or to be stored locally. In some cases however all the computed data is allowed through.

Perception-action hierarchy. The agent modules are organized as a *perception-action hierarchy*. This is an abstraction hierarchy, so that modules higher in the hierarchy process data of more abstract data types. We use a fixed number of levels of abstraction.

There are plans at different levels of abstraction, so a higher level planning module has a more abstract plan. The goal module has rules causing it to prioritize the set of goals that it has received, and to select the strongest one which is sent to the highest level plan module.

Dynamics. We devized a control system that tries all alternatives at each level until a *viable* plan and action are found. We defined a viable state as one that is driven by the current goal and is compatible with the currently perceived situation at all levels. This is achieved by selecting the strongest rule instance, sending it to the module below and waiting for a *confirmation* data item indicating that this datum caused activity in the module below. If a confirmation is not received within a given number of cycles then the rule instance is decremented for a given amount of time, allowing the next strongest rule instance to be selected, and so on.

A viable behavioral state corresponds to a coherent distributed process, with a selected dominant rule instance in each module, confirmed dynamically by confirmation signals from other modules.

3. Social Plans and Joint Action

We generalized the standard artificial intelligence representation of plan to one suitable for action by more than one collaborating agent. A *social plan* is a set of joint steps, with temporal and causal ordering constraints, each step specifying an action for every agent collaborating in the social plan, including the subject agent. The way an agent executes a plan is to attempt each joint step in turn. During a joint step it verifies that every collaborating agent is performing its corresponding action and then to attempt to execute its own corresponding individual action. We made most of the levels of the planning hierarchy work with social plans, the next to lowest works with a "selfplan" which specifies action only for the subject agent, and the lowest works with concrete motor actions. However, the action of these two lowest levels still depended on information received from the perception hierarchy.

Initial model and a social behavior. To make things more explicit, we'll now describe a simple joint behavior which is a prototype of many joint behaviors, namely the maintenance of affiliative relations in a group of agents by pairwise joint affiliative actions, usually called grooming.

The social relations module contained a long term memory of knowledge of affiliative relations among agents. This was knowledge of who is friendly with who and how friendly. This module kept track of affiliative actions and generated goals to affiliate with friends that had not been affiliated with lately. Each agent had stored social plans for grooming and for being groomed. Usually a subordinate agent with groom and dominant one will be groomed. We organized each social plan into four phases, as shown in the figure: orient, approach, prelude and groom, which could be evoked depending on the current state of the activities of the agents. Each phase corresponded to different rules being evoked.

Attention was controlled by the planning modules selecting the agents to participate with and communicating this choice to the higher levels of perception. These higher levels derived high level perceptual information only for those agents being attended to.

4. Autonomy, Situatedness and Voluntary Action

Autonomy. The concept of autonomy concerns the control relationship between the agent and other agents, including the user. As illustrated in our example, agents are autonomous, in the sense that they do not receive control imperatives and react to them, but instead each agent receives messages, and

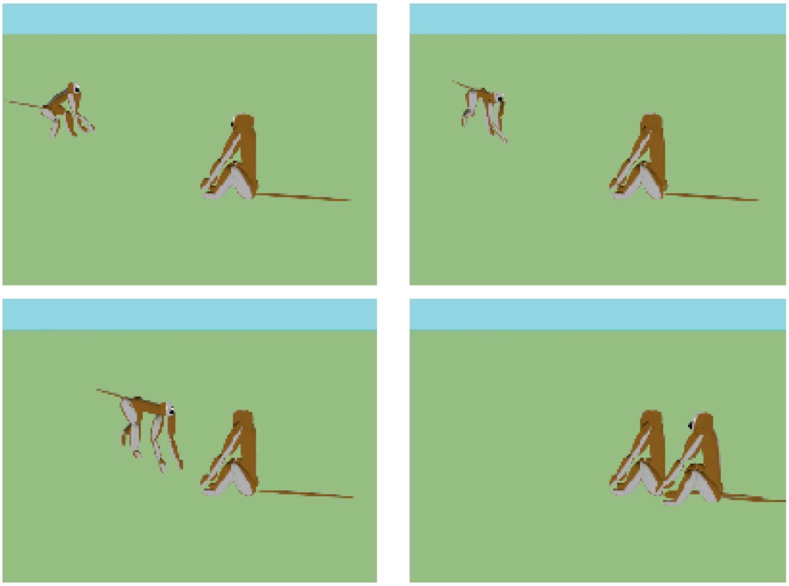


Figure 3.2. Four phases of grooming

perceives its environment, and makes decisions based on its own goals, and that is the only form of control for agents.

Further, agents may act continuously, and their behavior is not constrained to be synchronized with the user or other agents.

Constraint by commitments. A social agent is also constrained by any commitments it has made to other agents. In addition, we may have initially programmed it to be constrained by the general social commitments of the social group.

Voluntary control. The joint action is "voluntary" in the sense that each agent is controlled only by its own goals, plans and knowledge, and makes its own choices. These choices will be consistent with any commitments, and we are thus assuming that usually some choice exists after all such constraints are taken into account.

Situatedness of action. However the action of each agent is conditional upon what it perceives. If the external environment changes, the agent will change its behavior. This action is *situated* in the agent's external environment, to the extent that its decisions are dependent on or determined by this environment.

Thus, an agent is to some extent controlled by its environment. Environmental changes *cause* the agent to make different choices. If it rains, the agent will put its raincoat on, and if I stop the rain, the agent will take its raincoat off.

This assumes that the agent (i) does not make random and arbitrary actions, (ii) does not have a supersmart process which models everything and itself, in other words (iii) it is rational in the sense of using some not too complex reasoning or computational process to make its choices.

5. Mutual Planning and Control

Our agent architecture is flexibly both goal-directed and environmentally situated. It is also quite appropriate for social interaction, since the other agents are perceived at each level and can directly influence the action of the subject agent. It allows agents to enter into stable mutually controlled behaviors where each is perceived to be carrying out the requirements of the social plan of the other. Further, this mutually controlled activity is hierarchically organized, in the sense that control actions fall into a hierarchy of abstraction, from easily altered details to major changes in policy.

We implemented two kinds of social behavior, one was affiliation in which agents maintained occasional face-to-face interactions which boosted affiliation measures, and the other was social spacing in which agents attempted to maintain socially appropriate spatial relationships characterized by proximity, displacement and mutual observability. The set of agents formed a simple society which maintained its social relations by social action.

During an affiliation sequence, each of two interacting agents elaborates its selected social plan conditionally upon its perception of the other. In this way, both agents will scan possible choices until a course of action is found which is viable for both agents.

This constitutes mutual control. Note that the perception of the world by distal sensors is quite shared, however perception by tactile, proprioceptive, and visceral sensing is progressively more private and less shared. Each agent perceives both agents, which has some common and some private perception as input, and each agent executes its part of the joint action.

In each phase of grooming, each agent's social plan detects which phase it is in, has a set of expected perceptions of what the other may do, and a corresponding set of actions which are instantiated from the perception of what is actually perceived to occur. If, during a given phase, an agent changes its action to another acceptable variant within the same phase, then the other agent will simply perceive this and generate the corresponding action. If, on the other hand, one agent changes its action to another whose perception is not consistent with the other agent's social plan, then the other agent's social plan will fail at that level. In this latter case, rules will no longer fire at that level, so the level above will not receive confirmatory data and will start to scan for a viable plan at the higher level. This may result in recovery of the joint action without the first agent changing, however it is more likely that the induced change in the second agent's behavior will cause a similar failure and replanning activity in the first agent.

In the case of grooming, during orientation and approach, the groomee agent can move and also change posture, and the groomer will simply adjust, unless the groomee moves clearly away from the groomer, in which case the approach behavior will fail. When the groomer arrives at prelude distance, it expects the groomee to be not moving and to be looking at him, otherwise the prelude phase will not be activated. Then, if the groomee make a positive prelude response, the groomer can initiate the grooming phase.

Agents enter into, and terminate or modify, joint action voluntarily, each motivated by its own perceptions and goals.

6. Coparticipation and Engagement

Our notion of social plan has some subtlety and indirectness, which is really necessitated by the distributed nature of agent interaction. There is no agreed shared plan as such, each participant has their own social plan, which includes expectations of the actions of coparticipants. Each participant attempts to find and to carry out their "best" social plan which satisfies their goals. In constrained situations, it may be that the best social plan of each participant is very similar to the best social plans of coparticipants. Thus social plans of individuals may be more or less *engaged*. Engagement concens the agreement and coherence among the instantiations of the social plans of the participants.

A standard example is the prostitute and the client, which coparticipate and cooperate, each with his or her own goals and social plan. Thus, for social action, the prostitute needs to sufficiently match the client's social plan and model of prostitute appearance and behavior, and the client needs to behave sufficiently like the prostitute's idea of a client.

Adversarial coparticipation occurs with lawyers representing defendent and plaintiff. Since however there is always a residual conflict or disparity and residual shared benefits in all relationships, it is difficult to find cases of pure cooperation or even pure adversality.

The initiation (and termination) of joint action usually involves less engagement between the social plans of coparticipants. The grooming preludes observed in social monkeys are for example initially more unilateral. Initiation and termination usually involve protocols by which coparticipants navigate paths through a space of states of different degrees of engagement.

In this model, social interaction is never unilateral. First, some "other" is always an imagined coparticipant. Second, even in the case of hardwired evolved behaviors, the behavior is intended for, only works with, and only makes sense with, a coparticipant, even though, in this case, there is no explicit representation of the other. It is not clear for example what representation, if any, of the mother a baby may have. There is for example biological evidence of tuning of the babies sensory systems during pregnancy, and immediately after birth, to the mother's odor and voice. Thus, the mother constructs an explicit coparticipant and the baby acts as if it has a coparticipant.

7. Summary

We argued for and demonstrated an approach to social relationship, appropriate for agent-agent and user-agent interaction:

In a social relationship, agents enter into mutually controlled action regimes, which they maintain voluntarily by mutual perception and by the elaboration of their individual social plans.

Acknowledgement: This work was supported by the National Science Foundation, Information Technology and Organizations Program managed by Dr. Les Gasser, and is currently supported by the Cooperation and Social Systems program managed by Dr. Susan Iacono, Research grant IIS-9812714, and by the Caltech Neuromorphic Engineering Research Center, NSF Research Grant EEC-9730980.

References

[1] Alan H. Bond. Commitment: A Computational Model for Organizations of Cooperating Intelligent Agents. In *Proceedings of the 1990 Conference on Office Information Systems*, pages 21–30. Cambridge, MA, April 1990.

[2] Alan H. Bond. Describing Behavioral States using a System Model of the Primate Brain. *American Journal of Primatology*, 49:315–388, 1999.

[3] Alan H. Bond. Problem-solving behavior in a system model of the primate neocortex. *Neurocomputing*, to appear, 2001.

[4] Alan H. Bond and Les Gasser. An Analysis of Problems and Research in Distributed Artificial Intelligence. In *Readings in Distributed Artificial Intelligence*, pages 3–35. Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[5] Cindy Hazan and Debra Zeifman. Pair Bonds as Attachments. In Jude Cassidy and Phillip R. Shaver, editor, *Handbook of Attachment: Theory, Research and Clinical Applications*, pages 336–354. The Guilford Press, New York, 1999.

[6] L. C. Miller and S. J. Read. On the coherence of mental models of persons and relationships: a knowledge structure approach. In *Cognition in close relationships*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1991.

Chapter 4

DEVELOPING AGENTS WHO CAN RELATE TO US

Putting Agents in Our Loop via Situated Self-Creation

Bruce Edmonds
Centre for Policy Modelling, Manchester Metropolitan University

Abstract This paper addresses the problem of how to produce artificial agents so that they can relate to us. To achieve this it is argued that the agent must have humans in its developmental loop and not merely as designers. The suggestion is that an agent needs to construct its self as humans do - by adopting at a fundamental level others as its model for its self as well as vice versa. The beginnings of an architecture to achieve this is sketched. Some of the consequences of adopting such an approach to producing agents is discussed.

1. Introduction

In this paper I do not directly consider the question of how to make artificial agents so that humans can relate to them, but more the reverse: how to produce artificial agents so that they can relate to us. However, this is directly relevant to human-computer interaction since we, as humans, are used to dealing with entities who can relate to us - in other words, human relationships are reciprocal. The appearance of an ability in agents could allow a shift away from merely using them as tools towards forming relationships with them.

The basic idea is to put the human into the developmental loop of the agent so that the agent co-develops an identity that is intimately bound up with ours. This will give it a sound basis with which to base its dealings with us, enabling its perspective to be in harmony with our own in a way that would be impossible if one attempted to design such an empathetic sociality into it. The development of such an agent could be achieved by mimicking early human development in important respects - i.e. by socially situating it within a human culture.

The implementation details that follow derive from a speculative theory of the development of the human self that will be described. This may well be wrong but it seems clear that something of this ilk does occur in the development of young humans [23] [14]. So the following can be seen as simply a method to enable agents to develop the required abilities - other methods and processes may have the same effect.

2. The Inadequacy of the Design Stance for Implementing a Deeper Sociality

I (amongst others) have argued elsewhere that if an agent is to be embedded in its society (which is necessary if it is to have a part in the social constructs) then one will not be able to design the agent first and deploy it in its social context second, but rather that a considerable period of in situ *acculturation* will be necessary [10]. In addition to this it seems likely that several crucial aspects of the mind itself requires a society in order to develop, including intelligence [14] [13] and free-will [12].

Thus rather than specify directly the requisite social facilities and mechanisms I take the approach of specifying the social "hooks" needed by the agents and then evolve the social skills within the target society. In this way key aspects of the agent develop already embedded in the society which it will have to deal with. In this way the agent can truly partake of the culture around it. This directly mirrors the way our intelligence is thought to have evolved [18].

In particular I think that this process of embedding has to occur at an early stage of agent development for it to be most effective. In this paper I suggest that this needs to occur at an extremely basic stage: during the construction of the self. In this way the agent's own self will have been co-developed with its model of others and allow a deep empathy between agents and its society (in this case us).

3. A Model of Self Construction

Firstly I outline a model of how the self may be constructed in humans. This model attempts to reconcile the following requirements:

- That the self is only experienced indirectly [16].
- That a self requires a strong form of self-reference [20].
- That many aspects of the self are socially constructed [7].
- ”Recursive processing results from monitoring one's own speech” [5].

- That one has a ”narrative centre” [8].
- That there is a ”Language of Thought” [1] to the extent that high-level operations on the syntax of linguistic production, in effect, cause other actions.

The purpose of this model is to approach how we might provide the facilities for an agent to construct its self using social reflection via language use. Thus if the agent’s self is socially reflective this allows for a deep underlying commonality to exist without this needing to be prescribed beforehand. In this way the nature of the self can be develop within its society in a flexible manner and yet there be this structural commonality allowing empathy between its members. This model (of self development) is as follows:

1 There is a basic decision making process in the agents that acts upon the perceptions, actions and memories and returns decisions about new actions (that can include changing the focus of one’s perception and retrieving memories).

2 The agent does not have direct access to the workings of this basic process (i.e. it cannot directly introspect) but only of its perceptions and actions, past and present.

3 This basic process learns to choose its actions (including speech) to control its environment via its experiences (composed of its perceptions of its environment, its experiences of its own actions and its memories of both) including the other agents it can interact with. In particular it models the consequences of its actions (including speech acts). This basic mechanism produces primitive predictions (expectations) about the consequences of actions whose accuracy forms the basis for the learning mechanism. In other words the agent has started to make primitive models of its environment [4]. As part of this it also makes such model of other agents which it is ’pre-programmed’ to distinguish.

4 This process naturally picks up and tries out selections of the communications it receives from other agents and uses these as a basis (along with observed actions) for modelling the decisions of these other agents.

5 As a result it becomes adept at using communication acts to fulfil its own needs via others’ actions using its model of their decision making processes.

6 Using the language it produces itself it learns to model itself (i.e. to predict the decisions it will make) by applying its models of other agents to

itself by comparing its own and others' actions (including communicative acts). The richness of the language allows a relatively fine-grained transference of models of other's decision making processes onto itself.

7 Once it starts to model itself it quickly becomes good at this due to the high amount of direct data it has about itself. This model is primarily constructed in its language and so is accessible to introspection.

8 It refines its model of other agents using its self-model, attempting predictions of their actions based on what it thinks it would do in similar circumstances.

9 Simultaneously it refines its self-model from further observations of other's actions. Thus its model of other's and its own cognition co-evolve.

10 Since the model of its own decisions are made through language, it uses language production to implement a sort of high-level decision making process - this appears as a language of thought. The key points are that the basic decision making process are not experienced; the agent models others' decision making using their utterances as fine-grained indications of their mental states (including intentions etc.); and finally that the agent models itself by applying its model of others to itself (and vice versa). This seems to be broadly compatible with the summary of thinking on the language of thought [2].

4. General Consequences of this Model of Self Construction

The important consequences of this model are:

- The fact that models of other agent and self-models are co-developed means that many basic assumptions about one's own cognition can be safely projected to another's cognition and vice versa. This can form the basis for true empathetic relationships.

- The fact that an expressive language has allowed the modelling of others and then of its self means that there is a deep association of self-like cognition with this language.

- Communication has several sorts of use: as a direct action intended to accomplish some goal; as an indication of another's mental state/process; as an indication of one's own mental state/process; as an action designed to change another's mental state/process; as an action designed to change one's own mental state/process; etc.

- Although such agents do not have access to the basic decision making processes they do have access to and can report on their linguistic self-model which is a model of their decision making (which is, at least, fairly good). Thus, they do have a reportable language of thought, but one which is only a good approximation to the underlying basic decision making process.

- The model allows social and self reflective thinking, limited only by computational resources and ingenuity - there is no problem with unlimited regression, since introspection is done not directly but via a model of one's own thought processes.

5. Towards Implementing Self-Constructing Agents

The above model gives enough information to start to work towards an implementation. Some of the basic requirements for such an implementation are thus:

1. A suitable social environment (including humans)
2. Sufficiently rich communicative ability - i.e. a communicative language that allows the fine-grained modelling of others' internal states leading to action in that language
3. General anticipatory modelling capability
4. An ability to distinguish the experience of different types, including the observation of the actions of particular others; ones own actions; and other sensations
5. An ability to recognise other agents as distinguishable individuals
6. Need to predict other's decisions
7. Need to predict one's own decisions
8. Ability to reuse model structures learnt for one purpose for another

Some of these are requirements upon the internal architecture of an agent, and some upon the society it develops in. I will briefly outline a possibility for each. The agent will need to develop two sets of models.

1. A set of models that anticipate the results of action, including communicative actions (this roughly corresponds to a model of the world including other agents). Each model would be composed of several parts: - a condition for the action - the nature of the action - the anticipated effect of the action - (possibly) its past endorsements as to its past reliability

2 a set of candidate strategies for obtaining its goals (this roughly corresponding to plans); each strategy would also be composed of several parts: the goal; the sequence of actions, including branches dependent upon outcomes, loops etc.; (possibly) its past endorsements as to its past success.

These could be developed using a combination of anticipatory learning theory [15] as reported in [21] and evolutionary computation techniques. Thus rather than a process of inferring sub-goals, plans etc. they would be constructively learnt (similar to that in [9] and as suggested by [19]). The language of these models needs to be expressive, so that an open-ended model structure such as in genetic programming [17] is appropriate, with primitives to cover all appropriate actions and observations. Direct self-reference in the language to itself is not built-in, but the ability to construct labels to distinguish one's own conditions, perceptions and actions from those of others is important as well as the ability to give names to individuals. The language of communication needs to be a combinatorial one, one that can be combinatorially generated by the internal language and also deconstructed by the same.

The social situation of the agent needs to have a combination of complex cooperative and competitive pressures in it. The cooperation is necessary if communication is at all to be developed and the competitive element is necessary in order for it to be necessary to be able to predict other's actions [18]. The complexity of the cooperative/competitive mix encourages the prediction of one's own decisions. A suitable environment is where, in order to gain substantial reward, cooperation is necessary, but that inter-group competition occurs as well as competition for the dividing up of the rewards that are gained by a cooperative group.

Many of the elements of this model have already been implemented in pilot systems [9]; [11]; [21].

6. Consequences for Agent Production and Use

If we develop agents in this way, allowing them to learn their selves from within a human culture, we may have developed agents such that we can relate to them because they will be able to relate to us etc. The sort of social games which involve second guessing, lying, posturing, etc. will be accessible to the agent due to the fundamental empathy that is possible between agent and human. Such an agent would not be an 'alien' but (like some of the humans we relate to) all the more unsettling for that. To achieve this goal we will have to at least partially abandon the design stance and move more towards an enabling stance and accept the necessity of considerable acculturation of our agents within our society much as we do with our children.

7. Conclusion

If we want to put artificial agents truly into the "human-loop" then they will need to be able to reciprocate our ability to relate to them, including relating to them relating to us etc. In order to do this it is likely that the development of the agent's self-modelling will have to be co-developed with its modelling of the humans it interacts with. Just as our self-modelling has started to be influenced by our interaction with computers and robots [22], their self-modelling should be rooted in our abilities. One algorithm for this has been suggested which is backed up by a theory of the development of the human self. Others are possible. I argue elsewhere that if we carry on attempting a pure design stance with respect to the agents we create we will not be able to achieve an artificial intelligence (at least not one that would pass the Turing Test) [13]. In addition to this failure will be the lack of an ability to relate to us. Who would want to put anything, however sophisticated, in charge of any aspect of our life if it does not have the ability to truly relate to us - this ability is an essential requirement for many of the roles one might want agents for.

References

[1] Aydede. *Language of Thought Hypothesis: State of the Art*, 1999.
http://humanities.uchicago.edu/faculty/aydede/LOTH.SEP.html

[2] Aydede, M. and Güzeldere, G. Consciousness, Intentionality, and Intelligence: Some Foundational Issues for Artificial Intelligence. *Journal of Experimental and Theoretical Artificial Intelligence*, forthcoming.

[3] Barlow, H. The Social Role of Consciousness - Commentary on Bridgeman on Consciousness. *Psycoloquy* 3(19), *Consciousness* (4), 1992.

[4] Bickhard, M. H. and Terveen L. *Foundational Issues in Artificial Intelligence and Cognitive Science, Impasse and Solution*. New York: Elsevier Scientific, 1995.

[5] Bridgeman, B. On the Evolution of Consciousness and Language, *Psycoloquy* 3(15), Consciousness (1), 1992.

[6] Bridgeman, B. The Social Bootstrapping of Human Consciousness - Reply to Barlow on Bridgeman on Consciousness, *Psycoloquy* 3(20), *Consciousness* (5), 1992.

[7] Burns, T. R. and Engdahl, E. The Social Construction of Consciousness Part 2: Individual Selves, Self-Awareness, and Reflectivity. *Journal of Consciousness Studies*, 2:166-184, 1998.

[8] Dennett, D. C. *The Origin of Selves*, Cogito, 3:163-173, 1989.

[9] Drescher, G. L. *Made-up Minds, a Constructivist Approach to Artificial Intelligence*. Cambridge, MA: MIT Press, 1991.

[10] Edmonds, B. Social Embeddedness and Agent Development. *Proc. UKMAS'98*, Manchester, 1998.
http://www.cpm.mmu.ac.uk/cpmrep46.html

[11] Edmonds, B. Capturing Social Embeddedness: a Constructivist Approach. *Adaptive Behavior*, 7(3/4): 323-347, 1999.

[12] Edmonds, B. Towards Implementing Free-Will. AISB2000 Symposium on *How to Design a Functioning Mind*, Birmingham, 2000. http://www.cpm.mmu.ac.uk/cpmrep57.html

[13] Edmonds, B. The Constructability of Artificial Intelligence, *Journal of Logic, Language and Information*, 9:419-424, 2001.

[14] Edmonds, B. and Dautenhahn, K. The Contribution of Society to the Construction of Individual Intelligence. Workshop on *Socially Situated Intelligence*, SAB'98, Zürich, 1998. http://www.cpm.mmu.ac.uk:80/cpmrep42.html

[15] Hoffman, J. *Vorhersage und Erkenntnis* [Anticipation and Cognition]. Goettingen, Germany: Hogrefe, 1993.

[16] Gopnik, A. How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioural and Brain Sciences*, 16:1-14, 1993.

[17] Koza, J. R. *Genetic Programming: the programming of computers by means of natural selection*. Cambridge, MA: MIT Press, 1992.

[18] Kummer, H., Daston, L., Gigerenzer, G. and Silk, J. The social intelligence hypothesis. In Weingart et. al. (eds.), *Human by Nature: Between Biology and the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum, 157-179, 1997.

[19] Millgram, E. and Thagard, P. Deliberative Coherence. *Synthese*, 108(1):63-88, 1996.

[20] Perlis, D. Consciousness as Self-Function, *Journal of Consciousness Studies*, 4: 509-525, 1997.

[21] Stolzmann, W., Butz, M. V., Hoffmann, J. and Goldberg, D. E. First Cognitive Capabilities in the Anticipatory Classifier System. Proc. *Sixth International Conference on Simulation of Adaptive Behavior* (SAB 2000), MIT Press, 287-296, 2000.

[22] Turkle, S. *The Second Self, Computers and the Human Spirit*, New York: Simon and Schuster, 1984.

[23] Werner, E. The Ontogeny of the Social Self. Towards a Formal Computational Theory. In: Dautenhahn, K. (ed.) *Human Cognition and Social Agent Technology*, John Benjamins, 263-300, 1999.

Chapter 5

PARTY HOSTS AND TOUR GUIDES

Using Nonverbal Social Cues in the Design of Interface Agents to Support Human-Human Social Interaction

Katherine Isbister
Finali Corporation

Abstract Interface agents have the potential to be catalysts and orchestrators of human-human social interaction. To excel at this, agents must be designed to function well in a busy social environment, reacting to and conveying the kinds of primarily nonverbal social cues that help create and maintain the flow of social exchange.

This paper sets context for the sorts of cues that are important to track and to convey, and briefly describes two projects that incorporated such cues in agents that attempt to help the flow of human-human social interaction.

1. Introduction

1.1 The Importance of Nonverbal Social Cues

Nonverbal cues perform a variety of important functions in everyday human interaction, such as:

- Content and Mechanics: Nonverbal cues convey important content and conversational mechanics information, such as pointing out a location or setting up spatial relationships that complement what is said, indicating that one's turn is about to end, or setting a rhythm of emphasis (see Clark or Cassell for more comprehensive discussion of this topic).

- Social Intentions and Relationships: Nonverbal cues also express social intentions and interrelationships. For example, lovers will stand closer together than strangers; angry people may move closer to one another, turning up the proximity volume as they may turn up the volume of their voices (Hall).

- Attitudes: A good teacher indicates pride in the student through face and gesture (Lester et. al); a friendly nod indicates not just acceptance of an offer for coffee but enthusiasm toward that offer (Clark).

Nonverbal cues can include gestures made with the hands and head, expressions made with the face, posture, proximity, eye contact, as well as tone, volume, style, and duration of speech.

Nonverbal cues are routinely manipulated in human-human conversation to achieve certain goals, some admirable, some less so (Lester et. al point out the effectiveness of nonverbal cues in pedagogy; Cialdini notes that sales training often includes imitation of one's customer's body language, which increases that person's feeling of similarity to the salesperson, and thus likelihood of being convinced to buy).

1.2 Use of Nonverbal Social Cues in Interface Agents

There is experimental evidence confirming that people will also read nonverbal cues in agents, and that these nonverbal cues can in fact influence attitude toward the agent, as well as the level of behavioral influence the agent may have on the person (Isbister and Nass). Some examples of agents using nonverbal cues include:

- Deictic (content supporting) gestures in a virtual real estate agent (Bickmore and Cassell)

- Deictic and emotional gestures and facial expressions in a pedagogical agent (Lester et. al)

- Deictic, eye gaze, and turn-taking gestures in an agent meant to teach tasks within a shared virtual context (Rickel and Johnson).

Focus in these projects has been on the support of a one-on-one interaction with the agent.

1.3 Using Nonverbal Social Cues in Designing Interface Agents to Support Human-Human Communication

Agents with the ability to facilitate and enhance human-human social interaction could, for example, help to make connections between people with commonalities they do not yet know about, or guide group discovery and learning, among other potential applications.

In group settings, nonverbal cues are just as crucial as they are in one-on- one conversational settings. The same sorts of strategies apply, with some additional tactics related to group situations. For example, people use nonverbal cues to indicate when they are giving up or beginning a turn in a conversation (Clark), to welcome newcomers or ward off people who may be attempting to join a private conversation (Cassell), to indicate who they are referring to, or who might know more about a topic, and to help delineate conversational sub-groups within the main group (Clark, Hall).

To design a successful agent for this context, I believe there are several design factors to keep in mind:

- It's important that the agent 'knows' when to take the floor, and what value it might have when it does, as well as when to give up the floor.
- The agent should use proper turn-taking cues, and demonstrate sensitivity to facilitating the overall social flow of the conversation, rather than focussing on modelling or adapting to any one person.
- The agent should have a clear and appropriate social role, such as host or guide (see Isbister and Hayes-Roth for a demonstration of the effectiveness of an agent's social role in influencing visitor behavior).

In the sections that follow, I describe two interface agent projects which incorporated group-focused nonverbal social cue tracking and expression. *Please see the acknowledgements section of this paper for a list of contributors to this research.*

2. Helper Agent

2.1 Design of Helper Agent

Helper Agent supports human-human conversations in a video chat environment. Users have avatars they can move freely around the space, and Helper Agent is an animated, dog-faced avatar, which spends most of its time listening, at a distance. The agent tracks audio from two-person conversations, looking for longer silences. When it detects one, it approaches, directs a series of text-based, yes/no questions to both people, and uses their answers to guide its suggestion for a new topic to talk about. Then the agent retreats until needed again (see Figure 1).

Because Helper Agent is presented on-screen the same way users are, we could use nonverbal cues, such as turning to face users as it poses a question to them, and approaching and departing the conversation physically. The animations include nonverbal cues for asking questions, reacting to affirmative or negative responses, and making suggestions. The dog orients its face toward the user that it is addressing, with the proper expression for each phase: approach, first question, reaction, follow-up question, and finally topic suggestion. After concluding a suggestion cycle, the agent leaves the conversation zone, and meanders at a distance, until it detects another awkward silence. This makes it clear to the conversation pair that the agent need not be included in their discussion.

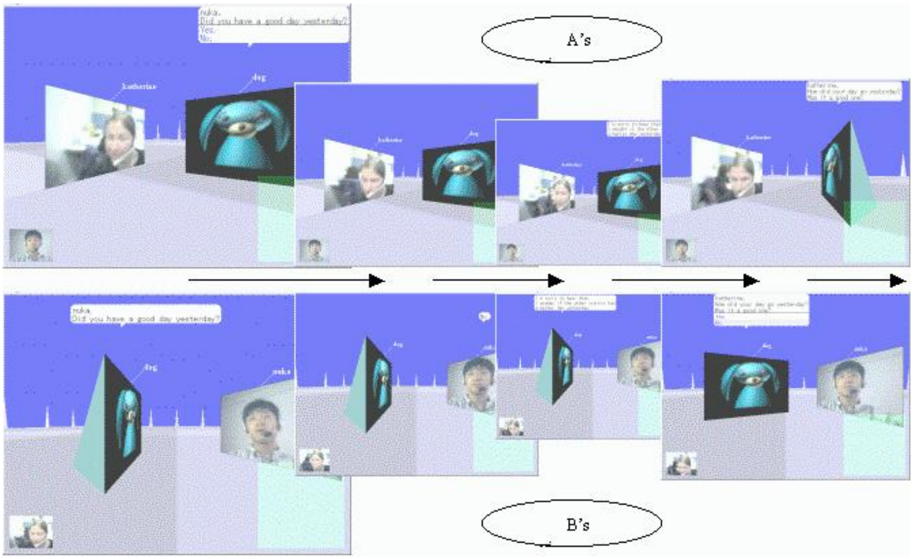

A's
nuka,
Did you have a good day yesterday?
Yes:
No:
Katherine
dog
Katherine,
How did your day go yesterday?
Was it a good one?
nuka
B's

Figure 5.1. Conversation from both participant's point-of-view: (1) person A is asked the first question (2) and responds, (3) then the agent comments. (4) Next person B is asked a question. Note that the agent faces the person it is addressing.

If the participants start talking again before the agent reaches them, it stops the approach and goes back to idling. The agent will also remain in idling state if the participants are standing far apart from each other (out of conversation range), or are not facing each other. If the participants turn away from each other during the agent's approach, or while it is talking, it will return to idling state, as well.

The agent decides there is silence when the sum of the voice volumes of both participants is below a fixed threshold value. When the agent detects a silence that lasts for more than a certain period of time, it decides the participants are in an awkward pause. The agent decides how to position itself, based on the location and orientation of each participant. The agent turns toward the participant that it's currently addressing. If the participants move while the agent is talking, the agent adjusts its location and orientation. The agent tries to pick a place where it can be seen well by both people, but also tries to avoid blocking the view between them. If it's hard to find an optimal position, the agent will stand so that it can at least be seen by the participant to whom it is addressing the question.

2.2 Evaluation of the Success of Helper Agent

We conducted an experiment to test the effectiveness of Helper Agent, in assisting in conversations between Japanese and American students. (For more about the method and results, please see Isbister, Nakanishi, Ishida, and Nass). People did engage with the agent. Most quickly grasped its purpose - accepting the agent as a valid participant, taking turns with it, and taking up its suggestions.

3. Tour Guide Agent

3.1 Designing Tour Guide Agent

The Tour Guide Agent project was part of Digital City Kyoto (http://www.digitalcity.gr.jp/). The tour was to be a point of entry to the online resource and to Kyoto, ideally increasing visitor interest in and use of the digital city. The tour was also designed to encourage dialogue and relationships among participants, and to increase exposure to Kyoto's history among friends and family of participants.

To create the agent's behavior, we observed tour guides, and read professional manuals on tour guide strategy (Pond). Strategies for storytelling that we imitated:

1. Stories were told about particular locations while in front of them.
2. Some stories included tales about previous tours.
3. Stories were selected partly because they were easy and fun to retell.
4. Guides adjusted timing and follow-up based on audience response.

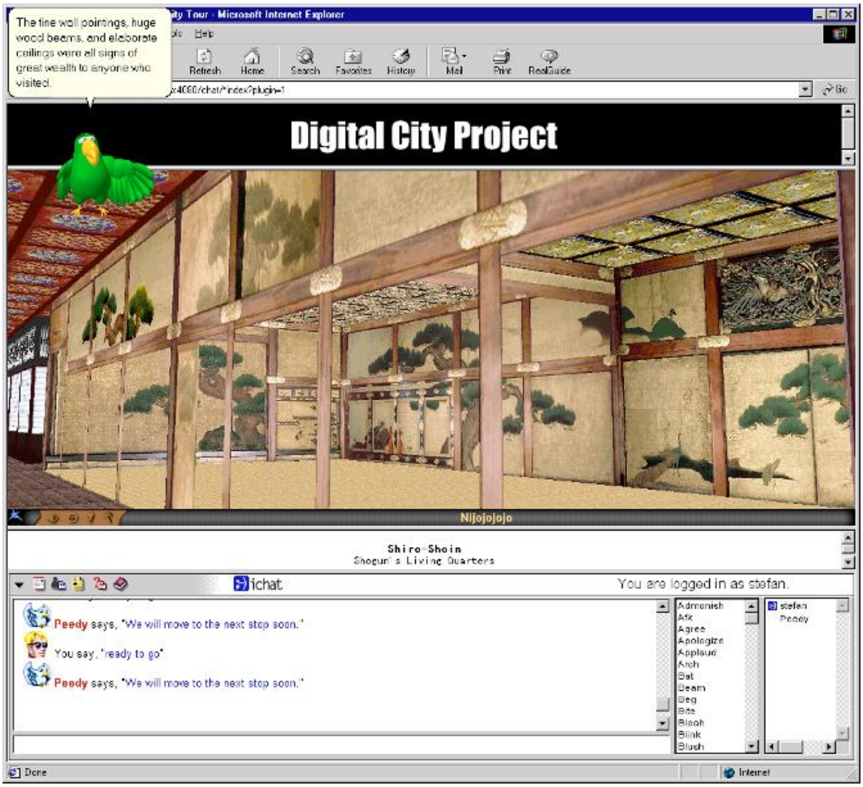
In our system, the digital tour-takers are all chatting in an online text environment, and use a simple 3-D control set to explore a virtual model of parts of Nijo Castle in Kyoto (see Figure 2). At each stop, the tour guide tells related stories, using gesture and expression to highlight key points.

The agent tracks the quantity of conversation, and looks for positive and negative keywords that indicate how visitors feel at the moment (negative words such as "boring, dull, too long"; positive words such as "wow, cool, neat, interesting"). The agent selects stories using a very simple decision rule (see Figure 3).

To make sure the tour stops for the right duration, the agent moves to the next stop only when a majority of tour-takers say they want to move forward. (For more about this project's technical details, please see Isbister).

3.2 Lessons Learned

Though we did not perform a formal evaluation, preliminary review of reactions to the tour indicated that the agent's stories were serving as a successful springboard for conversation, and worked nicely to supplement the visitors' experience of the virtual castle.



The fine wall paintings, huge wood beams, and elaborate ceilings were all signs of great wealth to anyone who visited.

Shiro-Shoin
Shogun's Living Quarters

You are logged in as stefan.

Peedy says, "We will move to the next stop soon."

You say, "ready to go"

Peedy says, "We will move to the next stop soon."

Figure 5.2. Kyoto Digital City Tour Guide Agent

	Valence of Conversation Contents	
Quantity of Talk	*Negative*	*Positive*
Low	medium length	long length
High	short length	medium length

Figure 5.3. Decision Rule for Agent Story Choice

4. Conclusions

In both agent projects, use of nonverbal social cues added value to human-human interaction. Of course, more exploration and evaluation is needed.

I encourage those in the social interface agent community to design agents for support roles such as tour guide or host, leaving the humans center stage. Design for group situations refocuses one's efforts to track and adapt to users, and creates an interesting new set of challenges. It also adds to the potentially useful applications for everyone's work in this field.

Acknowledgments

This research was conducted at NTT's Open Laboratory in Japan. Helper Agent was created with Hideyuki Nakanishi and Toru Ishida, of Kyoto University, and many other graduate and undergraduate students there. The study of Helper Agent was supported by Clifford Nass at Stanford University, as well as by Kyoto University. Tour Guide Agent was created with the engineering team at the NTT Open Lab, as well as students from Stanford University, resident at the Stanford Japan Center. Thanks to all for making this work possible.

References

[1] Edward T. Hall. *The Hidden Dimension.* Anchor Books, Doubleday, New York, 1982.

[2] Herbert H. Clark. *Using Language*. Cambridge University Press, Cambridge, England, 1996.

[3] James C. Lester and Stuart G. Towns and Charles B. Callaway and Jennifer L. Voerman and Patrick J. FitzGerald. Deictic and Emotive Communication in Animated Pedagogical Agents. In Cassell, Sullivan, Prevost, and Churchill, editor, *Embodied Conversational Agents*. M.I.T. Press, Cambridge, Massachusetts, 2000.

[4] Jeff Rickel and W. Lewis Johnson. Task-Oriented Collaboration with Embodied Agents in Virtual Worlds. In Cassell, Sullivan, Prevost, and Churchill, editor, *Embodied Conversational Agents*. M.I.T. Press, Cambridge, Massachusetts, 2000.

[5] Justine Cassell. Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In Cassell, Sullivan, Prevost, and Churchill, editor, *Embodied Conversational Agents*. M.I.T. Press, Cambridge, Massachusetts, 2000.

[6] Katherine Isbister. A Warm Cyber-Welcome: Using an Agent-Led Group Tour to Introduce Visitors to Kyoto. In T. Ishida and K. Isbister, editor, *Digital Cities: Technologies, Experiences, And Future Perspectives*. Springer-Verlag, Berlin, 1998.

[7] Katherine Isbister and Barbara Hayes-Roth. Social Implications of Using Synthetic Characters, in . In *Animated Interface Agents: Making Them Intelligent (a workshop in IJCAI-97, Nagoya, JAPAN, August 1997)*, pages 19–20. 1997.

[8] Katherine Isbister and Clifford Nass. Consistency of Personality in Interactive Characters: Verbal Cues, Non-verbal Cues, and User Characteristics. *International Journal of Human Computer Studies*, 2000.

[9] Katherine Isbister and Hideyuki Nakanishi and Toru Ishida and Clifford Nass. Helper Agent: Designing an Assistant for Human-Human Interaction in a Virtual Meeting Space. In *Proceedings CHI 2000 Conference*, 2000.

[10] Katherine Pond. *The Professional Guide: Dynamics of Tour Guiding*. Van Nostrand Reinhold Co., New York, 1993.

[11] Robert B. Cialdini. *Influence: The Psychology of Persusasion*. Quill, William Morrow, New York, 1984.

[12] Timothy Bickmore and Justine Cassell. Relational Agents: A Model and Implementation of Building User Trust. In *Proceedings CHI 2001 Conference*, 2001.

Chapter 6

INCREASING SIA ARCHITECTURE REALISM BY MODELING AND ADAPTING TO AFFECT AND PERSONALITY

Eva Hudlicka
Psychometrix Associates, Inc.

Abstract The ability to exhibit, recognize and respond to different affective states is a key aspect of social interaction. To enhance their believability and realism, socially intelligent agent architectures must be capable of modeling and generating behavior variations due to distinct affective states on the one hand, and to recognize and adapt to such variations in the human user / collaborator on the other. This chapter describes an adaptive user interface system capable of recognizing and adapting to the user's affective and belief state: the Affect and Belief Adaptive Interface System (ABAIS). ABAIS architecture implements a four-phase adaptive methodology and provides a generic adaptive framework for exploring a variety of user affect assessment methods and GUI adaptation strategies. An ABAIS prototype was implemented and demonstrated in the context of an Air Force combat task, using a knowledge-based approach to assess and adapt to the pilot's anxiety level.

1. Introduction

A key aspect of human-human social interaction is the ability to exhibit and recognize variations in behavior due to different affective states and personalities. These subtle, often non-verbal, behavioral variations communicate critical information necessary for effective social interaction and collaboration. To enhance their believability and realism, socially intelligent agent architectures must be capable of modeling and generating behavior variations due to distinct affective states and personality traits on the one hand, and to recognize and adapt to such variations in the human user / collaborator on the other. We have been pursuing these goals along two lines of research: (1) developing a cognitive architecture capable of modeling a variety of individual differences (e.g., affective states, personality traits, etc.) [5], and (2) developing an adaptive user interface capable of recognizing and adapting to the user's affective and belief state (e.g., heightened level of anxiety, belief in imminent threat, etc.) [4].

In this chapter we focus on the area of affective adaptation and describe an Affect and Belief Adaptive Interface System (ABAIS) designed to compensate for performance biases caused by users' affective states and active beliefs. The performance bias prediction is based on empirical findings from emotion research, and knowledge of specific task requirements. The ABAIS architecture implements a four-phase adaptive methodology: (1) assessing user affect and belief state; (2) identifying their potential impact on performance; (3) selecting a compensatory strategy; and (4) implementing this strategy in terms of specific GUI adaptations. ABAIS provides a generic adaptive framework for exploring a variety of user assessment methods (e.g., knowledge-based, self-reports, diagnostic tasks, physiological sensing), and GUI adaptation strategies (e.g., content- and format-based). We outline the motivating psychological theory and empirical data, and present preliminary results from an initial prototype implementation in the context of an Air Force combat task. We conclude with a summary and outline of future research and potential applications for the synergistic application of the affect-adaptive and affect and personality modeling methodologies within SIA architectures.

2. Selecting Affective States And Personality Traits

The first step for both the modeling and the adaptation research goals is to identify key affective and personality traits influencing behavior. The *affective states* studied most extensively include anxiety, positive and negative affect, and anger. The effects of these states range from influences on distinct information processes (e.g., attention and working memory capacity, accuracy, and speed; memory recall biases), through autonomic nervous system manifestations (e.g., heart rate, GSR), to visible behavior (e.g., facial expressions, approach vs. avoidance tendencies, etc.) [9, 7, 1]. A wide variety of *personality traits* have been studied, ranging from general, abstract behavioral tendencies such as the Five Factor Model or "Big 5" (Extraversion, Emotional Stability, Agreeableness, Openness, Conscientiousness) and "Giant 3" (Approach behaviors, Inhibition behaviors, Aggressiveness) personality traits, through psychodynamic / clinical traits (e.g., narcissistic, passive-aggressive, avoidant, etc.), to characteristics relevant for particular type of interaction (e.g., style of leadership, etc.) [3, 8]. Our initial primary focus in both the modeling and the adaptation research areas was on anxiety, aggressiveness, and obsessiveness.

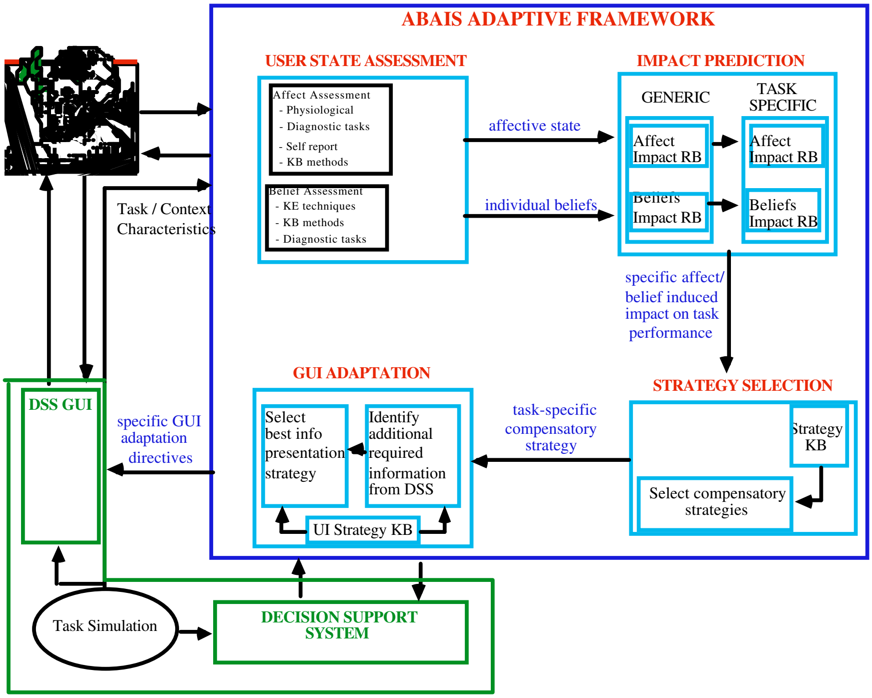


Figure 6.1. ABAIS Affect-Adaptive Architecture.

3. Adaptive Methodology and Architecture

We developed a *methodology* designed to compensate for performance biases caused by users' affective states and active beliefs [4]. The methodology consists of four stages: 1) assessing the user's affective state and performance-relevant beliefs; 2) identifying their potential impact on performance (e.g., focus on threatening stimuli); 3) selecting a compensatory strategy (e.g., presentation of additional information to reduce ambiguity); and 4) implementing this strategy in terms of specific GUI adaptations (e.g., presenting additional information, or changing information format to enhance situation awareness). This methodology was implemented within an architecture: the Affect and Belief Adaptive Interface System (ABAIS). The ABAIS *architecture* consists of four modules, described below, each implementing the corresponding step of the adaptive methodology (see Figure 6.1): User State Assessment, Impact Prediction, Strategy Selection, and GUI Adaptation.

User State Assessment Module. This module receives a variety of data about the user and the task context, and from these data identifies the user's predominant affective state (e.g., high level of anxiety) and situation-relevant beliefs (e.g., interpretation of ambiguous radar return as threat), and their potential influence on task performance (e.g., firing a missile). Since no single reliable method currently exists for affective assessment, the User Assessment module provides facilities for the flexible combination of multiple methods. These include: physiological assessment (e.g., heart rate); diagnostic tasks; self-reports; and use of knowledge-based methods to derive likely affective state based on factors from current task context (e.g., type, complexity, time of day, length of task), personality (negative emotionality, aggressiveness, obsessiveness, etc.), and individual history (past failures and successes, affective state associated with current task, etc.). For the preliminary ABAIS prototype, we focused on a knowledge-based assessment approach, applied to assessment of anxiety levels, to demonstrate the feasibility of the overall adaptive methodology. The knowledge-based assessment approach assumes the existence of multiple types of data (e.g., individual history, personality, task context, physiological signals), and from these data derives the likely anxiety level. Anxiety was selected both because it is the most prevalent affect during crisis situations, and because its influence on cognition has been extensively studied and empirical data exist to support specific impact prediction and adaptation strategies.

Impact Prediction Module. This module receives as input the identified affective states and associated task-relevant beliefs, and determines their most likely influence on task performance. The goal of the impact prediction module is to predict the influence of a particular affective state (e.g., high anxiety) or belief state (e.g., "aircraft under attack", "hostile aircraft approaching", etc.) on task performance. Impact prediction process uses rule-based reasoning (RBR) and takes place in two stages. First, the *generic effects* of the identified affective state are identified, using a knowledge-base that encodes empirical evidence about the influence of specific affective states on cognition and performance. Next, these generic effects are instantiated in the context of the current task to identify *task-specific effects*, in terms of relevant domain entities and procedures (e.g., task prioritization, threat assessment). The knowledge encoded in these rules is derived from a detailed cognitive affective personality task analysis (CAPTA), which predicts the effects of different affective states and personality traits on performance in the current task context. The CAPTA process is described in detail in [6]. The separation of the generic and specific knowledge enhances modularity and simplifies knowledge-based adjustments.

Table 6.1. Examples of Task-Specific Rules for Strategy Selection.

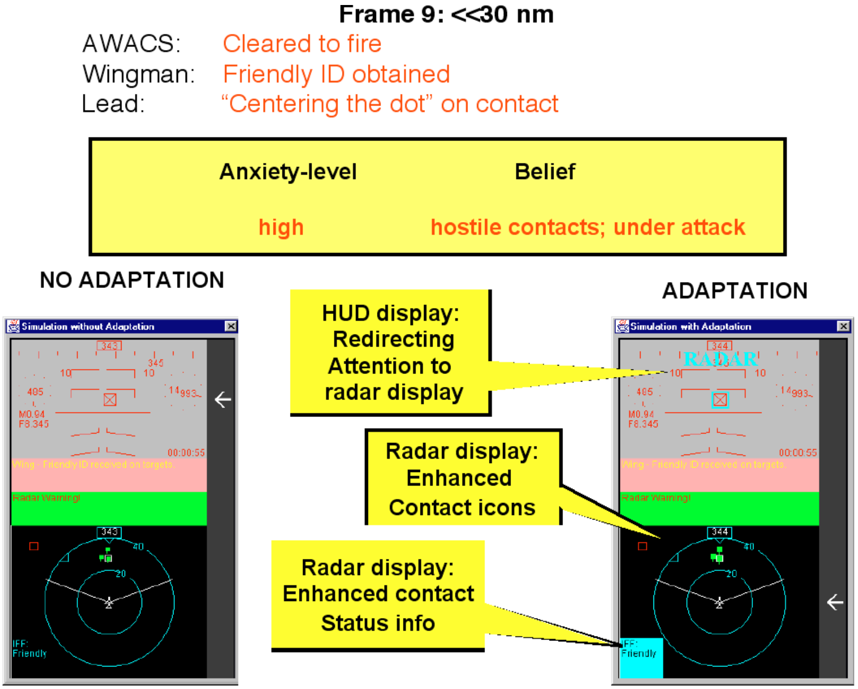
Anxiety effects
IF (recent change in radar return status) THEN (emphasize change in status)
IF (attention focus = HUD) AND (incoming radar data) THEN (redirect focus to radar)
IF (attention focus = radar) AND (Incoming radio call) THEN (redirect focus to radio)
IF (likelihood of task neglect for <instrument> = high) & (has-critical-info? <instrument>) THEN (emphasize <instrument> visibility)
IF (target = unknown) AND (target belief = hostile) THEN (emphasize unknown status) AND (collect more data)
Aggressiveness effects
IF (likelihood of premature attack = high) THEN (display all available info about enemy a/c) AND(enhance display of enemy a/c info)
Obsessiveness effects
IF (likelihood of delayed attack = high) THEN (display all available enemy a/c info) AND (display likelihood of enemy attack) AND (display vulnerability envelope) AND (display reminders for attack tasks)

Strategy Selection Module. This module receives as input the predicted specific effects of the affective and belief states, and selects a compensatory strategy to counteract resulting performance biases. Strategy selection is accomplished by rule-based reasoning, where the rules map specific performance biases identified by the Impact Prediction Module (e.g., task neglect, threat-estimation bias, failure-estimation bias, etc.) onto the associated compensatory strategies (e.g., present reminders of neglected tasks, present broader evidence to counteract threat-estimation bias, present contrary evidence to counteract failure-driven confirmation bias, etc.). Table 6.1 shows examples of task-specific rules for compensatory strategy selection.

GUI Adaptation Module. This module performs the final step of the adaptive methodology, by implementing the selected compensatory strategy in terms of specific GUI modifications. A rule-based approach is used to encode the knowledge required to map the specific compensatory strategies onto the necessary GUI adaptations. The specific GUI modifications take into consideration information about the individual pilot preferences for information presentation, encoded in customized user preference profiles; for example, highlighting preferences might include blinking vs. color change vs. size change of the relevant display or icon. In general, two broad categories of adaptation are possible: content-based, which *provide additional information*, and format-based, which *modify the format of existing information* (see Figure 6.2).

4. Results

The ABAIS prototype was implemented and demonstrated in the context of an Air Force combat mission, used a knowledge-based approach to assess



Frame 9: <<30 nm

AWACS: Cleared to fire
Wingman: Friendly ID obtained
Lead: "Centering the dot" on contact

Anxiety-level	Belief
high	hostile contacts; under attack

NO ADAPTATION

ADAPTATION

HUD display: Redirecting Attention to radar display

Radar display: Enhanced Contact icons

Radar display: Enhanced contact Status info

Simulation without Adaptation
Simulation with Adaptation
RADAR
Wing - Friendly ID received on targets
Radar Warning!
IFF: Friendly

Figure 6.2. Example of Specific Scenario Adaptation Sequence.

the pilot's anxiety level, and modified selected cockpit instrument displays in response to detected increases in anxiety levels. Several representative pilot profiles were defined, varying in personality, physiological responsiveness, training, individual history, and adaptation preferences, making it more or less likely that the pilot would reach a specific level of anxiety during the mission. Once an increased level of anxiety was observed, ABAIS predicted that the heightened level of anxiety would cause narrowing of attention, an increased focus on potentially threatening stimuli, and a perceptual bias to interpret ambiguous radar signals as threats, thus risking fratricide. ABAIS therefore suggested a compensatory strategy aimed at: 1) directing the pilot's attention to a cockpit display showing the recent status change; and 2) enhancing the relevant signals on the radar to improve detection. Figure 6.2 illustrates these adaptations (not easily visible in a black and white version of the figure). Specifically, the blinking, enlarged, blue contact icon on the HUD display indicates a change in status. A blinking blue "RADAR" string displayed on the HUD, the pilot's current focus, directs the pilot to look at the radar display, which shows an enhanced contact icon indicating a change in status, with details provided in the text box in lower left corner of the display.

5. Conclusions

We described a research area aimed at producing more realistic behavior in socially intelligent agents, namely the *recognition of, and adaptation to, a user's affective state*. We developed an adaptive methodology and demonstrated its effectiveness by implementing a prototype Affect and Belief Adaptive Interface System (ABAIS). ABAIS assessed the user affect and belief states using a knowledge-based approach and information from a variety of sources, predicted the effects of this state within the constrained context of the demonstration task, and suggested and implemented specific GUI adaptation strategies based on the pilot's individual information presentation preferences. The preliminary results indicate the general feasibility of the approach, raise a number of further research questions, and provide information about the specific requirements for a successful, operational affective adaptive interface. Although the initial prototype was developed within a military domain, we believe that the results are applicable to a broad variety of non-military application areas, as outlined below.

Requirements for Adaptation. A number of requirements were identified as necessary for affective adaptive interface system implementation. These include: Limiting the number, type, and resolution of affective states, and using multiple methods and data sources for affective state assessment; providing individualized user data and user-customized knowledge-bases; implementing 'benign' adaptations that at best enhance and at worst maintain current level of situation awareness (i.e., never limit access to existing information).

Key Issues to Address. A number of issues must be addressed to further validate this approach and to provide robust affect adaptive systems. These include: an empirical evaluation; multiple-method affect assessment; and demonstration of the ABAIS methodology across multiple task contexts.

Future Work. Possible future work in the broad area of user affect and modeling is limitless at this point, as the endeavor is in its infancy. Key questions include issues such as: What emotions *should* and *can* be addressed in adaptive systems? When should an agent attempt to enhance the user's affective state, adapt to the user's affective state, or attempt to counteract it? Cañamero offers an excellent summary of some of the affect-related issues that must be addressed by the SIA architecture research community [2].

Individually, both modeling and recognition of affective and beliefs states, and personality traits provide a powerful enhancement to agent architectures, by enabling socially intelligent, adaptive behavior. The coordinated integration of these two enhancements within a single agent architecture promises even further benefits, by enhancing the realism and effectiveness of human-machine interaction across a variety of application areas, including education and training, virtual reality assessment and treatment environments, and real-time decision aids in crisis-prone contexts.

Acknowledgments

The research described in this chapter was supported in part by US Air Force Contract F41624–98–C5032. We would like to acknowledge the contributions of the following individuals to this project: the technical monitor Dr. Michael McNeese of Air Force Research Laboratory, Mr. John Billingsley of Psychometrix Associates, Inc., Mr. Bob Shaw of NTI, Inc., Dr. Thomas von Wiegand and Prof. Rosalind Picard of MIT, and Dr. Greg Zacharias of Charles River Analytics, Inc.

References

[1] J.T. Cacioppo, D.J. Klein, G.G. Bernston, and E. Hatfield. The Psychophysiology of Emotion. In M. Lewis and J. Haviland, editors, *Handbook of Emotions*. Guilford Press, New York, 1993.

[2] L.D. Cañamero. Issues in the Design of Emotional Agents. In *Emotional and Intelligent: The Tangled Knot of Cognition. Papers from the 1998 AAAI Fall Symposium*. TR FS-98–03, pages 49–54. AAAI Press, Menlo Park, CA, 1998.

[3] P.T. Costa and R.R. McCrae. Four ways five factors are basic. *Personality and Individual Differences*, 13: 653–665, 1992.

[4] E. Hudlicka and J. Billingsley. *ABAIS: Affect and Belief Adaptive Interface System*. Report AFRL-HE-WP-TR-1999–0169. WPAFB, OH: US AFRL, 1999.

[5] E. Hudlicka and J. Billingsley. Representing Behavior Moderators in Military Human Performance Models. In *Proceedings of the 8th Conference on Computer Generated Forces and Behavioral Representation*, pages 423–433. Orlando, FL, May 1999.

[6] E. Hudlicka. Cognitive Affective Personality Task Analysis. Technical Report 0104, Psychometrix Associates, Inc., Blacksburg, VA, 2001.

[7] J.E. LeDoux. Cognitive-Emotional Interactions in the Brain. *Cognition and Emotion*, 3(4): 267–289, 1989.

[8] G. Matthews and I.J. Deary. *Personality Traits*. Cambridge University Press, Cambridge, UK, 1998.

[9] J.M.G. Williams, F.N. Watts, C. MacLeod, and A. Mathews. *Cognitive Psychology and Emotional Disorders*. John Wiley, New York, 1997.

Chapter 7

COOPERATIVE INTERFACE AGENTS

Sebastiano Pizzutilo, Berardina De Carolis and Fiorella de Rosis
Department of Computer Science, University of Bari

Abstract Animated agents are endowed with personality and emotions, with the aim of increasing their believability and of establishing an "empathetic" relationship with the user. In this chapter, we claim that, to endow agents with social intelligence, the communication traits described in the Five Factor Model should be integrated with some cooperation attitudes. We describe our experience in building an agent that combines the two personality aspects and discuss the problems still open.

1. Introduction

In the near future, computers will either "disappear", to ubiquitously pervade life environment in a not immediately perceivable way, or take the appearance of a human being, to undertake a friendship relationship with the user. In both cases, endowing agents with some form of social intelligence appears to be a crucial need. If reasoning, help and control abilities are distributed among specialised agents integrated with objects of daily life, some form of communication and cooperation among them will be needed, to avoid conflicting behaviours. Moreover, a "believable" embodied agent should be able to understand the users, help them in solving problems, find ways of coming to a mediated solution in case of conflicts and so on. To enable the users to foresee how the agent will behave, it should harmonise its external appearance with its internal behaviour, understand how to adapt to their needs and moods and, finally, enable them to "select a different partner" if they wish.

Short and long-term variations in the behaviour of embodied agents have been metaphorically represented, respectively, in terms of *emotional states* and *personality traits*. Endowing socially intelligent agents with a personality requires defining: (1) which forms of social intelligence these agents should have and how they may be translated in terms of personality traits; (2) how a trait may be represented in the agent's mental state and reasoning style; (3) how various traits may be combined in the same individual; and finally (4) how one or more traits may be manifested in the agent's behaviour. In this chapter we discuss our experience in building an Interface Agent that cooperates with the user in performing software application tasks; we will focus our description on the way that we formalised its cooperation attitude.

2. Cooperation personalities in XDM-Agent

Research on personality-based Human-Computer Interaction (HCI) has be driven by results of studies about human intelligence—in particular, the Five-Factor Model (FFM) and the Interpersonal Circumplex Model (IC). The FFM [10] derives from the psychologists' need of defining "the most important ways in which individuals differ in their enduring emotional, interpersonal, experiential, attitudinal and motivational styles" [10]. The five dimensions (Extraversion, Emotional Stability, Agreeableness, Openness, and Conscientiousness) are an interpretation of results of applying factor analysis to questionnaires submitted to various groups of subjects; their meaning is a subjective interpretation of the set of variables they "explain", and is described with natural language terms. "Sociability" or "Social closeness" is associated, in particular, with Extraversion. The second method employed to categorise human personalities is Wiggins' measure of IC [13], with axes "Dominance" and "Affiliation". Whether the two factorisation criteria are related is not fully clear: some authors identify Extraversion with Dominance, while others argue that Extraversion is best seen as located midway between Dominance and Warmth [10].

Researchers in HCI have employed the two mentioned factorisation criteria to enrich interfaces with a personality. Some notable examples: Nass and colleagues studied graphical interfaces in terms of Dominance and agent-based interfaces in terms of Extraversion [11]; Dryer adopted the IC model [8]; André and colleagues [1] attach Extraversion and Agreeableness to the members of their "Presentation Teams"; Ball and Breese [3] included Dominance and Friendliness in their modelling of personality-related observable behaviour. To computer scientists, the advantage of referring to the two mentioned models is to have a widely accepted frame of reference, with a definition of the way that every personality factor manifests itself in the external behaviour. The main disadvantage is that these personality traits refer to a characterisation of communication styles rather than to mental social attitudes. They are therefore very useful for endowing agents with a "pleasant" and "believable" appearance, but not to express diversification in social relationships. Another difficulty in employing the cited models is that traits are defined through natural language descriptions and are not easily formalised into the "mental state" of an agent. The first and most relevant contribution to a cognitive theory of personalities was due to Carbonell [4], who saw them as combinations of degrees of importance assigned to goals. A second example, to which we will refer in particular in this chapter, is Castelfranchi and Falcone's theory of cooperation in multi-agent systems [5].

Although affective expressions may contribute to increase interface agents' friendliness, its acceptability is driven by the level of help provided to the user, that is by its "cooperation attitude". This *level of help* should not be equal for all users but should be tailored to their attitudes towards computers in general, and towards the specific software to which the agent is applied in particular. These attitudes may be synthesised in a *level of delegation* of tasks that the user adopts towards the agent. To select the helping attitude that best suits the user needs, the agent has to be endowed with a reasoning capacity that enables it to observe the user, to model her expected abilities and needs and to plan the "best" response in every context. We had already applied the theory of Castelfranchi and Falcone to formalise the mental state of agents and their reasoning capacities in our a Project GOLEM [6]. With the project described in this chapter, we extend that research in the direction of embodied animated agents.

XDM-Agent is an embodied animated character that helps the user in performing the tasks of a given application; its cooperation attitude changes according to the user and the context. Although the agent is domain-independent, we will take electronic mail as a case study, to show some examples of how it behaves in helping to use Eudora. In a software of wide use like this, all procedures should be very natural and easy to perform. The first goal of XDM-Agent is then "to make sure that the user performs the main tasks without too much effort". At the same time, the agent should avoid providing too much help when this is not needed; a second goal is therefore "to make sure that the user does not see the agent as too intrusive or annoying". These general goals may specialise into more specific ones, according to the "cooperation attitude" of the agent. In deciding the level and the type of help to provide, XDM-Agent should consider, at the same time, the user experience and her "delegation attitude". The agent's decision of whether and how to help the user relies on the following knowledge sources:

Own Mental State. This is the representation of the agent's goals (Goal XDM (T g)) and abilities (Bel XDM (CanDo XDM a)) and the actions it intends to perform (Bel XDM (IntToDo XDM a)).

Domain Knowledge. XDM should know all the plans that enable achieving tasks in the application: $\forall g \forall p$ (Domain-Goal $g) \wedge$(Domain-Plan $p) \wedge$(Achieves p $g) \Rightarrow$ (KnowAbout XDM $g) \wedge$ (KnowAbout XDM $p) \wedge$ (Know XDM (Achieves p g)). It should know, as well, the individual steps of every domain-plan: $\forall g \forall a$ (Domain-Goal $p) \wedge$(Domain-action $a) \wedge$ (Step a $p) \Rightarrow$ (KnowAbout XDM $p) \wedge$ (KnowAbout XDM $a) \wedge$ (Know XDM (Step a p)).

User Model. The agent should have some hypothesis about: (1) the user goals, both in general and in specific phases of interaction [$\forall g$ (Goal U $(T\ g)) \Rightarrow$ (Bel XDM (Goal U (Tg)))]; (2) her abilities [$\forall a$ (CanDo U $a) \Rightarrow$(Bel XDM (CanDo U a))]; and (3) what the user expects the agent to do, in every phase of interaction [$\forall a$ (Goal U (IntToDo XDM a)) $\Rightarrow$(Bel XDM Goal U (IntToDo XDM a))]. This may be default, stereotypical knowledge about the user that is settled at the beginning of the interaction. Ideally, the model should be updated dynamically, through plan recognition.

Reasoning Rules. The agent employs this knowledge to take decisions about the level of help to provide in any phase of interaction, according to its helping attitude, which is represented as a set of reasoning rules. For instance, if XDM-Agent is a *benevolent*, it will respond to all the user's (implicit or explicit) requests of performing actions that it presumes she is not able to do:

Rule R1 $\forall a$[(Bel XDM (Goal U (IntToDo XDM a)))$\wedge$(Bel XDM $\neg$ (CanDo U a))$\wedge$(Bel XDM (CanDo XDM a))] $\Rightarrow$(Bel XDM (IntToDo XDM a)).

If, on the contrary, the agent is a *supplier*, it will do the requested action only if this does not conflict with its own goals:

Rule R2 $\forall a$ [(Bel XDM (Goal U (IntToDo XDM a)))$\wedge$ (Bel XDM (CanDo XDM a)) $\wedge$ ($\neg\exists$ g (Goal XDM (T g) $\wedge$ (Bel XDM (Conflicts a g)))] $\Rightarrow$ (Bel XDM (IntToDo XDM a))

. . . and so on for the other personality traits.

Let us assume that our agent is benevolent and that the domain goal g is to write a correct email address. In deciding whether to help the user, it will have to check, first of all, how the goal g may be achieved. Let us assume that no conflict exists between g and the agent's goals. By applying rule R1, XDM will come to the decision to do its best to help the user in writing the address, by directly performing all the steps of the plan. The agent might select, instead, a *level of help* to provide to the user; this level of help may be seen, as well, as a personality trait. If, for instance, XDM-Agent is a *literal helper*, it will only check that the address is correct. If, on the contrary, it is an *overhelper*, it will go beyond the user request of help to hypothesize her higher-order goal (for instance, to be helped in correcting the address, if possible). A *subhelper* will only send a generic error message; this is what Eudora does at present if the user tries to send a message without specifying any address. If, finally, the user asks the agent to suggest how to correct the string and the agent is not able to perform this action and is a *critical helper*, it will select and apply, instead, another plan it knows.

3. Personality Traits' Combination

In multiagent cooperation, an agent may find itself in the position of delegating some task or helping other agents. A theory is therefore needed to establish how delegation and helping attitudes may combine in the same agent. Some general thoughts about this topic may be found in [6]. In XDM-Agent, the agent's reasoning on whether to help the user ends up with an intentional state—to perform an individual action, an entire plan or part of a plan. This intentional state is transformed into an action that may include communication with the user; for instance, an *overhelper* agent will interact with the user to specify the error included in the string, will propose alternatives on how the string might be corrected and will ask the user to correct it. In this phase, the agent will adopt a communication personality trait—for instance, it might do it in an "extroverted" or an "introverted" way. The question then is *how should cooperation and communication personalities be combined?* Is it more reasonable to assume that an overhelper is extroverted or introverted? We do not have, at present, an answer to this question. In the present prototype, we implemented only two personalities (a *benevolent* and a *supplier*) and we associated the benevolent trait with the extroverted one and the supplier with the introverted.

The user's desire to receive help may be formalised, as well, in personality terms. If the user is a *lazy*, she expects to receive, from XDM, some cooperation in completing a task, even if she would be able to do it by herself (and therefore, irrespectively of her level of experience):

Rule R3 $\forall a \forall g$[(Goal U (T g))$\wedge$(Bel U (Achieves a g))$\wedge$ (Bel XDM (CanDo XDM a)) $\Rightarrow$ (Goal U (IntToDo XDM a))].

If, on the contrary, the user is a *delegating-if-needed*, she will need help only if she is not able to do the job by herself (for instance, if she is a novice):

Rule R4 $\forall a \forall g$ [(Goal U (T g))$\wedge$(Bel U (Achieves a g))$\wedge$(Bel XDM $\neg$ (CanDo U a))$\wedge$(Bel XDM (CanDo XDM a)) $\Rightarrow$(Goal U (IntToDo XDM a))].

Providing help to an expert and "delegating-if-needed" user will be seen as a kind of intrusiveness that will violate the agent's goal to avoid annoying the user.

In our first prototype of XDM-Agent, the agent's cooperation personality (and therefore its helping behaviour) may be settled by the user at the beginning of the interaction or may be selected according to some hypothesis about the user. As we said before, the agent should be endowed with a plan recognition ability that enables it to update dynamically its image of the user. Notice that, while recognising communication traits requires observing the external (verbal and nonverbal) behaviour of the user, inferring the cooperation attitude requires reasoning on the history of interaction (a cognitive diagnosis task that we studied, in probabilistic terms, in [7]). Once some hypothesis about the user's delegation personality exists, how should the agent's helping personality be settled? One of the controversial results of research about communication personalities in HCI is whether the similarity or the complementarity principles hold—that is, whether an "extroverted" interface agent should be proposed to an "extroverted" user, or the contrary. When cooperation personalities are considered, the question becomes the following: How much should an interface agent help a user? How much importance should be given to the user experience (and therefore her abilities in performing a given task), and how much to her propensity to delegate that task? In our opinion, the answer to this question is not unique. If XDM-Agent's *goals* are those mentioned before, that is "to make sure that the user performs the main tasks without too much effort" and "to make sure that the user does not see the agent as too much intrusive or annoying", then the following combination rules may be adopted:

CR1 (DelegatingIfNeeded U) ⇒ (Benevolent XDM): The agent helps delegating-if-needed users only if it presumes that they cannot do the action by themselves.

CR2 (Lazy U) ⇒ (Supplier XDM): The agent does its best to help lazy users, unless this conflicts with its own goals.

. . . and so on. However, if the agent has also the goal *to make sure that users exercise their abilities* (such as in Tutoring Systems), then the matching criteria will be different; for instance:

CR3 (Lazy U) ⇒(Benevolent XDM): The agent helps a lazy user only after checking that she is not able to do the job by herself. In this case, the agent's cooperation behaviour will be combined with a communication behaviour (for instance, Agreeableness) that warmly encourages the user in trying to solve the problem by herself.

XDM-Agent has been implemented by trying to achieve a distinction between its external appearance (its "Body", developed with MS-Agent) and its internal behaviour (its "Mind", developed in Java). It appears as a character that can take several bodies, can move on the display to indicate objects and make several other gestures, can speak and write a text in a balloon. To ensure that its body is consistent with its mind, the ideal would be to match the agent's appearance with its helping personality; however, as we said, no data are available on how cooperation traits manifest themselves, while literature is rich on how communication traits are externalised. At present, therefore, XDM-Agent's body only depends on its communication personality. We associate a different character with each of them (*Genie* with the benevolent-extroverted and *Robby* with the supplier-introverted). However, MS-Agent enables us to program the agent to perform a minimal part of the gestures we would need. We are therefore working, at the same time, to develop a more refined animated agent that can adapt its face, mouth and gaze to its high-level goals, beliefs and emotional states. This will enable us to directly link individual components of the agent's mind to its verbal and non-verbal behaviour, through a set of personality-related activation rules [12].

4. Conclusions

Animated agents tend to be endowed with a personality and with the possibility to feel and display emotions, for several reasons. In Tutoring Systems, the display of emotions enables the agent to show to the students that it cares about them and is sensitive to their emotions; it helps convey enthusiasm and contributes to ensure that the student enjoys learning [9]. In Information-Providing Systems, personality traits contribute to specify a motivational profile of the agent and to orient the dialog accordingly [1]. Personality and emotions are attached to Personal Service Assistants to better "anthropomorphize" them [2]. As we said at the beginning of this chapter, personality traits that are attached to agents reproduce the "Big-Five" factors that seem to characterise human social relations. Among the traits that have been considered so far, "Dominance/Submissiveness" is the only one that relates to cooperation attitudes. According to Nass and colleagues, "Dominants" are those who pretend that others help them when they need it; at the same time, they tend to help others by assuming responsibilities on themselves. "Submissives", on the contrary, tend to obey to orders and to delegate actions and responsibilities whenever possible. This model seems, however, to consider only some combinations of cooperation and communication attitudes that need to be studied and modelled separately and more in depth. We claim that Castelfranchi and Falcone's theory of cooperation might contribute to such a goal, and the first results obtained with our XDM-Agent prototype encourage us to go on in this direction. As we said, however, much work has still to be done to understand how psychologically plausible configurations of traits may be defined, how they evolve dynamically during interaction, and how they are externalised.

References

[1] E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. The Automated Design of Believable Dialogues for Animated Presentation Teams. In J. Cassel, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 220–255. The MIT Press, Cambridge, MA, 2000.

[2] Y. Arafa, P. Charlton, A. Mamdani, and P. Fehin. Designing and Building Personal Service Assistants with Personality. In S. Prevost and E. Churchill, editors, *Proceedings of the Workshop on Embodied Conversational Characters*, pages 95–104, Tahoe City, USA, October 12–15, 1998.

[3] G. Ball and J. Breese. Emotion and Personality in a Conversational Agent. In J. Cassel, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 189–219. The MIT Press, Cambridge, MA, 2000.

[4] J. Carbonell. Towards a Process Model of Human Personality Traits. *Artificial Intelligence*, 15: 49–74, 1980.

[5] C. Castelfranchi and R. Falcone. Towards a Theory of Delegation for Agent-Based Systems. *Robotics and Autonomous Systems*, 24(3/4): 141–157, 1998.

[6] C. Castelfranchi, F. de Rosis, R. Falcone, and S. Pizzutilo. Personality Traits and Social Attitudes in Multiagent Cooperation. *Applied Artificial Intelligence*, 12: 7–8, 1998.

[7] F. de Rosis, E. Covino, R. Falcone, and C. Castelfranchi. Bayesian Cognitive Diagnosis in Believable Multiagent Systems. In M.A. Williams and H. Rott, editors, *Frontiers of Belief Revision*, pages 409–428. Kluwer Academic Publisher, Applied Logic Series, Dordrecht, 2001.

[8] D.C. Dryer (1998). Dominance and Valence: A Two-Factor Model for Emotion in HCI. In *Emotional and Intelligent: The Tangled Knot of Cognition. Papers from the 1998 AAAI Fall Symposium.* TR FS-98–03, pages 76–81. AAAI Press, Menlo Park, CA, 1998.

[9] C. Elliott, J.C. Lester, and J. Rickel. Interpreting Affective Computing into Animated Tutoring Agents. In *Proceedings of the 1997 IJCAI Workshop on Intelligent Interface Agents: Making Them Intelligent*, pages 113–121. Nagoya, Japan, August 25, 1997.

[10] R.R. McCrae and O. John, O. An Introduction to the Five-Factor Model and its Applications. *Journal of Personality*, 60: 175–215, 1992.

[11] C. Nass, Y. Moon, B.J. Fogg, B. Reeves, and D.C. Dryer. Can Computer Personalities Be Human Personalities? *International Journal of Human-Computer Studies*, 43: 223–239, 1995.

[12] I. Poggi, C. Pelachaud, and F. de Rosis. Eye Communication in A Conversational 3D Synthetic Agent. *AI Communications*, 13(3): 169–181, 2000.

[13] J.S. Wiggins and R. Broughton. The Interpersonal Circle: A Structural Model for the Integration of Personality Research. *Perspective in Personality*, 1: 1–47, 1985.

Chapter 8

PLAYING THE EMOTION GAME WITH FEELIX

What Can a LEGO Robot Tell Us about Emotion?

Lola D. Cañamero
Department of Computer Science, University of Hertfordshire

Abstract This chapter reports the motivations and choices underlying the design of Feelix, a simple humanoid LEGO robot that displays different emotions through facial expression in response to physical contact. It concludes by discussing what this simple technology can tell us about emotional expression and interaction.

1. Introduction

It is increasingly acknowledged that social robots and other artifacts interacting with humans must incorporate some capabilities to express and elicit emotions in order to achieve interactions that are natural and believable to the human side of the loop. The complexity with which these emotional capabilities are modeled varies in different projects, depending on the intended purpose and richness of the interactions. Simple models have for example been integrated in affective educational toys for small children [7], or in robots performing a particular task in very specific contexts [11]. Sophisticated robots designed to entertain socially rich relationships with humans [1] incorporate more complex and expressive models. Finally, other projects such as [10] have focused on the study of emotional expression for the sole purpose of social interaction; this was also our purpose in building Feelix[1]. We approached this issue from a "minimalist" perspective, using a small set of features that would make emotional expression and interaction believable and at the same time easily analyzable, and that would allow us to assess to what extent we could rely on the tendency humans have to anthropomorphize in their interactions with objects presenting human-like features [8].

Previous work by Jakob Fredslund on Elektra[2], the predecessor of Feelix, showed that: (a) although people found it very natural to interpret the happy and sad expressions of Elektra's smiley-like face, more expressions were needed to engage them in more interesting and long-lasting interactions; and (b) a clear causal pattern for emotion elicitation was necessary for people to attribute intentionality to the robot and to "understand" its displays. We turned to psychology as a source of inspiration for more principled models of emotion to design Feelix. However, we limited our model in two important ways. First, expression (and its recognition) was restricted to the face, excluding other elements that convey important emotion-related information such as speech or body posture. Since we wanted Feelix's emotions to be clearly recognizable, we opted for a category approach rather than for a componential (dimensional) one, as one of the main criteria used to define emotions as basic is their having distinctive prototypical facial expressions. Second, exploiting the potential that robots offer for physical manipulation—a very primary and natural form of interaction—we restricted interaction with Feelix to tactile stimulation, rather than to other sensory modalities that do not involve physical contact.

What could a very simple robot embodying these ideas tell us about emotional expression and interaction? To answer this question, we performed emotion recognition tests and observed people spontaneously playing with Feelix.

2. Feelix

Due to space limitations, we give below a very general description of the robot and its emotion model, and refer the reader to [3] for technical details.

2.1 The Robot

Feelix is a 70cm-tall "humanoid" robot (Figure 8.1) built from commercial LEGO Mindstorms™ robotic construction kits. Feelix expresses emotions by means of its face. To interact with the robot, people sit or stand in front of it. Since we wanted the interaction to be as natural as possible, the feet seemed the best location for tactile stimulation, as they are protruding and easy to touch; we thus attached a binary touch sensor underneath each foot.

Feelix's face has four degrees of freedom (DoF) controlled by five motors, and makes different emotional expressions by means of two eyebrows (1 DoF) and two lips (3 DoF). The robot is controlled on-board by two LEGO Mindstorms RCX™ computers[3], which communicate via infrared messages.

2.2 Emotion Model

Feelix can display the subset of basic expressions proposed by Ekman in [4], with the exception of disgust—i.e. anger, fear, happiness, sadness, and surprise, plus a neutral face[4]. Although it is possible to combine two expressions in Feelix's face, the robot has only been tested using a winner-take-all

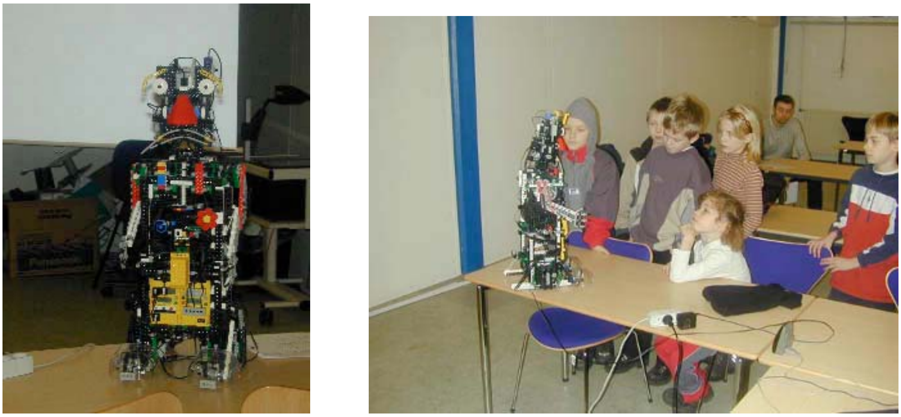


Figure 8.1. Left: Full-body view of Feelix. Right: Children guessing Feelix's expressions.

strategy[5] based on the level of emotion activation to select and display the emotional state of the robot.

To define the "primitives" for each expression we have adopted the features concerning positions of eyebrows and lips usually found in the literature, which can be described in terms of Action Units (AUs) using the Facial Action Coding System [6]. However, the constraints imposed by the robot's design and technology (see [3]) do not permit the exact reproduction of the AUs involved in all of the expressions (e.g., inner brows cannot be raised in Feelix); in those cases, we adopted the best possible approximation to them, given our constraints. Feelix's face is thus much closer to a caricature than to a realistic model of a human face.

To elicit Feelix's emotions through tactile stimulation, we have adopted the generic model postulated by Tomkins [12], which proposes three variants of a single principle: (1) A sudden increase in the level of stimulation can activate both positive (e.g., interest) and negative (e.g., startle, fear) emotions; (2) a sustained high level of stimulation (overstimulation) activates negative emotions such as distress or anger; and (3) a sudden stimulation decrease following a high stimulation level only activates positive emotions such as joy. We have complemented Tomkins' model with two more principles drawn from a homeostatic regulation approach to cover two cases that the original model did not account for: (4) A low stimulation level sustained over time produces negative emotions such as sadness (understimulation); and (5) a moderate stimulation level produces positive emotions such as happiness (well-being). Feelix's emotions, activated by tactile stimulation on the feet, are assigned different intensities calculated on the grounds of stimulation patterns designed on the above principles. To distinguish between different kinds of stimuli using only binary touch sensors, we measure the *duration* and *frequency* of the presses applied to the feet. The type of stimuli are calculated on the basis of a minimal time unit or *chunk*. When a chunk ends, information about stimuli—their number and type—is analyzed and the different emotions are assigned intensity levels according to the various stimulation patterns in our emotion activation model. The emotion with the highest intensity defines the emotional state and expression of the robot. This model of emotion activation is implemented by means of a timed finite state machine described in [3].

3. Playing with Feelix

Two aspects of Feelix's emotions have been investigated: the understandability of its facial expressions, and the suitability of the interaction patterns.

Emotion recognition tests[6], detailed in [3], are based on subjects' judgments of emotions expressed by faces, both in movement (the robot's face) and still (pictures of humans). Our results are congruent with findings about recognition of human emotional expressions reported in the literature (e.g., [5]). They show that the "core" basic emotions of anger, happiness, and sadness are most easily recognized, whereas fear was mostly interpreted as anxiety, sadness, or surprise. This latter result also confirms studies of emotion recognition from pictures of human faces, and we believe it might be due to structural similarities among those emotional expressions (i.e. shared AUs) or/and to the need of additional expressive features. Interestingly, children were better than adults at recognizing emotional expressions in Feelix's caricaturized face when they could freely describe the emotion they observed, whereas they performed worse when given a list of descriptors to choose from. Contrary to our initial guess, providing a list of descriptors diminished recognition performance for most emotions both in adults and in children.

The plausibility of the interactions with Feelix has been informally assessed by observing and interviewing the same people spontaneously interacting with the robot. Some activation patterns (those of happiness and sadness) seem to be very natural and easy to understand, while others present more difficulty (e.g., it takes more time to learn to distinguish between the patterns that activate surprise and fear, and between those that produce fear and anger). Some interesting "mimicry" and "empathy" phenomena were also found. In people trying to elicit an emotion from Feelix, we observed their mirroring—in their own faces and in the way they pressed the feet—the emotion they wanted to elicit (e.g., displaying an angry face and pressing the feet with much strength while trying to elicit anger). We have also observed people reproducing Feelix's facial expressions during emotion recognition, this time with the reported purpose of using proprioception of facial muscle position to assess the emotion observed. During recognition also, people very often mimicked Feelix's expression with vocal inflection and facial expression while commenting on the expression ('ooh, poor you!', 'look, now it's happy!'). People thus seem to "empathize" with the robot quite naturally.

4. What Features, What Interactions?

What level of complexity must the emotional expressions of a robot have to be better recognized and accepted by humans? The answer partly depends on the kinds of interactions that the human-robot couple will have. The literature, mostly about analytic models of emotion, does not provide much guidance to the designer of artifacts. Intuitively, one would think that artifacts inspired by a category approach have simpler designs, whereas those based on a componential approach permit richer expressions. For this purpose, however, more complex is not necessarily better, and some projects, such as [10] and Feelix, follow the idea put forward by Masahiro Mori (reported, e.g., in [9]) that the progression from a non-realistic to a realistic representation of a living thing is nonlinear, reaching an "uncanny valley" when similarity becomes almost, but not quite perfect[7]; a caricaturized representation of a face can thus be more acceptable and believable to humans than a realistic one, which can present distracting elements for emotion recognition and where subtle imperfections can be very disturbing. Interestingly, Breazeal's robot Kismet [1], a testbed to investigate infant-caretaker interactions, and Feelix implement "opposite" models based on dimensions and categories, respectively, opening up the door to an investigation of this issue from a synthetic perspective. For example, it would be very interesting to investigate whether Feelix's expressions would be similarly understood if designed using a componential perspective, and to single out the meaning attributed to different expressive units and their roles in the emotional expressions in which they appear. Conversely, one could ask whether Kismet's emotional expression system could be simpler and based on discrete emotion categories, and still achieve the rich interactions it aims at.

Let us now discuss some of our design choices in the light of the relevant design guidelines proposed by Breazeal in [2] for robots to achieve human-like interaction with humans.

Issue I. *The robot should have a cute face to trigger the 'baby-scheme' and motivate people to interact with it.* Although one can question the cuteness of Feelix, the robot does present some of the features that trigger the 'baby-scheme'[8], such as a big head, big round eyes, and short legs. However, none of these features is used in Feelix to express or elicit emotions. Interestingly, many people found that Feelix's big round (fixed) eyes were disturbing for emotion recognition, as they distracted attention from the relevant (moving) features. In fact, it was mostly Feelix's expressive behavior that elicited the baby-scheme reaction.

Issue II. *The robot's face needs several degrees of freedom to have a variety of different expressions, which must be understood by most people.* The insufficient DoF of Elektra's face was one of our motivations to build Feelix. The question, however, is how many DoF are necessary to achieve a particular kind of interaction. Kismet's complex model, drawn from a componential approach, allows to form a much wider range of expressions; however, not all of them are likely to convey a clear emotional meaning to the human. On the other hand, we think that Feelix's "prototypical" expressions associated to a discrete emotional state (or to a combination of two of them) allow for easier emotion recognition—although of a more limited set—and association of a particular interaction with the emotion it elicits. This model also facilitates an incremental, systematic study of what features are relevant (and how) to express or elicit different emotions. Indeed, our experiments showed that our features were insufficient to express fear, were body posture (e.g., the position of the neck) adds much information.

Issue IV. *The robot must convey intentionality to bootstrap meaningful social exchanges with the human.* The need for people to perceive intentionality in the robot's displays was another motivation underlying the design of Feelix's emotion model. It is however questionable that "more complexity" conveys "more intentionality" and adds believability, as put forward by the uncanny valley hypothesis. As we observed with Feelix, very simple features can have humans put much on their side and anthropomorphize very easily.

Issue V. *The robot needs regulatory responses so that it can avoid interactions that are either too intense or not intense enough.* Although many behavioral elements can be used for this, in our robot emotional expression itself acted as the only regulatory mechanism influencing people's behavior—in particular sadness as a response to lack of interaction, and anger as a response to overstimulation.

5. Discussion

What can a LEGO robot tell us about emotion? Many things, indeed. Let us briefly examine some of them.

Simplicity. First, it tells us that for modeling emotions and their expressions simple is good . . . but not when it is too simple. Building a highly expressive face with many features can be immediately rewarding as the attention it is likely to attract from people can lead to very rich interactions; however, it might be more difficult to evaluate the significance of those features in eliciting humans' reactions. On the contrary, a minimalist, incremental design approach that starts with a minimal set of "core" features allows us not only to identify more easily what is essential[9] versus unimportant, but also to detect missing features and flaws in the model, as occurred with Feelix's fear expression.

Beyond surface. Second, previous work with Elektra showed that expressive features alone are not enough to engage humans in prolonged interaction. Humans want to understand expressive behavior as the result of some underlying causality or intentionality. Believability and human acceptance can only be properly achieved if expressive behavior responds to some clear model of emotion activation, such as tactile stimulation patterns in our case.

Anthropomorphism. Feelix also illustrates how, as far as emotion design is concerned, realism and anthropomorphism are not always necessary . . . nor necessarily good. Anthropomorphism is readily ascribed by the human partner if the robot has the right features to trigger it. The designer can thus rely to some extent on this human tendency, and build an emotional artifact that can be easily attributed human-like characteristics. Finding out what makes this possible is, in our opinion, an exciting research challenge. However, making anthropomorphism an essential part of the robot's design might easily have the negative consequences of users' frustrated expectations and lack of credibility.

Multidisciplinarity. Finally, it calls for the need for multidisciplinary collaboration and mutual feedback between researchers of human and artificial emotions. Feelix implements two models of emotional interaction and expression inspired by psychological theories about emotions in humans. This makes Feelix not only very suitable for entertainment purposes, but also a proof-of-concept that these theories can be used within a synthetic approach that complements the analytic perspective for which they were conceived. We do not claim that our work provides evidence regarding the scientific validity of these theories, as this is out of our scope. We believe, however, that expressive robots can be very valuable tools to help human emotion researchers test and compare their theories, carry out experiments, and in general think in different ways about issues relevant to emotion and emotional/social interactions.

Acknowledgments

I am indebted to Jakob Fredslund for generously adapting his robot Elektra to build Feelix and for helping program the robot and perform the tests, and to Henrik Lund for making this research possible. Support was provided by the LEGO-Lab, Department of Computer Science, University of Aarhus, Denmark.

Notes

1. FEELIX: FEEL, Interact, eXpress.

2. www.daimi.au.dk/∼chili/elektra.html.

3. One RCX controls the emotional state of the robot on the grounds of tactile stimulation applied to the feet, while the other controls its facial displays.

4. Visit www.daimi.au.dk/∼chili/feelix/feelix_home.htm for a video of Feelix's basic expressions.

5. I have also built some demos where Feelix shows chimerical expressions that combine an emotion in the upper part of the face—eyebrows—and a different one in the lower part—mouth.

6. Tests were performed by 86 subjects—41 children, aged 9–10, and 45 adults, aged 15–57. All children and most adults were Danish. Adults were university students and staff unfamiliar with the project, and visitors to the lab.

7. I am grateful to Mark Scheeff for pointing me to this idea, and to Hideki Kozima for helping me track it down. Additional information can be found at www.arclight.net/∼pdb/glimpses/valley.html.

8. According to Irenäus Eibl-Eibesfeldt, the baby-scheme is an "innate" response to treat as an infant every object showing certain features present in children. See for example I. Eibl-Eibesfeldt, *El hombre preprogramado*, Alianza Universidad, Madrid, 1983 (4th edition); original German title: *Der vorprogrammierte Mensch*, Verlag Fritz Molden, Wien-München-Zürich, 1973.

9. As an example, the speed at which the expression is formed was perceived as particularly significant in sadness and surprise, especially in the motion of eyebrows.

References

[1] C. Breazeal. Designing Sociable Machines: Lessons Learned. *This volume*.

[2] C. Breazeal and A. Forrest. Schmoozing with Robots: Exploring the Boundary of the Original Wireless Network. In K. Cox, B. Gorayska, and J. Marsh, editors, *Proc. 3rd. International Cognitive Technology Conference*, pages 375–390. San Francisco, CA, August 11–14, 1999.

[3] L.D. Cañamero and J. Fredslund. I Show You How I Like You—Can You Read It in my Face? *IEEE Trans. on Systems, Man, and Cybernetics: Part A*, 31(5): 454–459, 2001.

[4] P. Ekman. An Argument for Basic Emotions. *Cognition and Emotion*, 6(3/4): 169–200, 1992.

[5] P. Ekman. Facial Expressions. In T. Dalgleish and M. Power, editors, *Handbook of Cognition and Emotion*, pages 301–320. John Wiley & Sons, Sussex, UK, 1999.

[6] P. Ekman and W.V. Friesen. *Facial Action Coding System.* Consulting Psychology Press, Palo Alto, CA, 1976.

[7] D. Kirsch. *The Affective Tigger: A Study on the Construction of an Emotionally Reactive Toy*. S.M. thesis, Department of Media Arts and Sciences, Massachusetts Institute of Technology, Cambridge, MA, 1999.

[8] B. Reeves and C. Nass. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press/CSLI Publications, New York, 1996.

[9] J. Reichard. *Robots: Fact, Fiction + Prediction*. Thames & Hudson Ltd., London, 1978.

[10] M. Scheeff, J. Pinto, K. Rahardja, S. Snibbe and R. Tow. Experiences with Sparky, a Social Robot. *This volume*.

[11] S. Thrun. Spontaneous, Short-term Interaction with Mobile Robots in Public Places. In *Proc. IEEE Intl. Conf. on Robotics and Automation*. Detroit, Michigan, May 10–15, 1999.

[12] S.S. Tomkins. Affect Theory. In K.R. Scherer and P. Ekman, editors, *Approaches to Emotion*, pages 163–195. Lawrence Erlbaum, Hillsdale, NJ, 1984.

Chapter 9

CREATING EMOTION RECOGNITION AGENTS FOR SPEECH SIGNAL

Valery A. Petrushin
Accenture Technology Labs

Abstract This chapter presents agents for emotion recognition in speech and their application to a real world problem. The agents can recognize five emotional states—unemotional, happiness, anger, sadness, and fear—with good accuracy, and be adapted to a particular environment depending on parameters of speech signal and the number of target emotions. A practical application has been developed using an agent that is able to analyze telephone quality speech signal and to distinguish between two emotional states—"agitation" and "calm". This agent has been used as a part of a decision support system for prioritizing voice messages and assigning a proper human agent to respond the message at a call center.

1. Introduction

This study explores how well both people and computers can recognize emotions in speech, and how to build and apply emotion recognition agents for solving practical problems. The first monograph on expression of emotions in animals and humans was written by Charles Darwin in the 19th century [4]. After this milestone work psychologists have gradually accumulated knowledge in this field. A new wave of interest has recently risen attracting both psychologists and artificial intelligence (AI) specialists. There are several reasons for this renewed interest such as: technological progress in recording, storing, and processing audio and visual information; the development of non-intrusive sensors; the advent of wearable computers; the urge to enrich human-computer interface from point-and-click to sense-and-feel; and the invasion on our computers of life-like agents and in our homes of robotic animal-like devices like Tiger's Furbies and Sony's Aibo, which are supposed to be able express, have and understand emotions [6]. A new field of research in AI known as *affective computing* has recently been identified [10]. As to research on recognizing emotions in speech, on one hand, psychologists have done many experiments and suggested theories (reviews of about 60 years of research can be found in [2, 11]). On the other hand, AI researchers have made contributions in the following areas: emotional speech synthesis [3, 9], recognition of emotions [5], and using agents for decoding and expressing emotions [12].

2. Motivation

The project is motivated by the question of how recognition of emotions in speech could be used for business. A potential application is the detection of the emotional state in telephone call center conversations, and providing feedback to an operator or a supervisor for monitoring purposes. Another application is sorting voice mail messages according to the emotions expressed by the caller.

Given this orientation, for this study we solicited data from people who are not professional actors or actresses. We have focused on negative emotions like anger, sadness and fear. We have targeted telephone quality speech (less than 3.4 kHz) and relied on voice signal only. This means that we have excluded modern speech recognition techniques. There are several reasons to do this. First, in speech recognition emotions are considered as noise that decreases the accuracy of recognition. Second, although it is true that some words and phrases are correlated with particular emotions, the situation usually is much more complex and the same word or phrase can express the whole spectrum of emotions. Third, speech recognition techniques require much better quality of signal and computational power.

To achieve our objectives we decided to proceed in two stages: research and development. The objectives of the first stage are to learn how well people recognize emotions in speech, to find out which features of speech signal could be useful for emotion recognition, and to explore different mathematical models for creating reliable recognizers. The second stage objective is to create a real-time recognizer for call center applications.

3. Research

For the first stage we had to create and evaluate a corpus of emotional data, evaluate the performance of people, and select data for machine learning. We decided to use high quality speech data for this stage.

3.1 Corpus of Emotional Data

We asked thirty of our colleagues to record the following four short sentences: “This is not what I expected”, “I’ll be right there”, “Tomorrow is my birthday”, and “I’m getting married next week.” Each sentence was recorded by every subject five times; each time, the subject portrayed one of the following emotional states: happiness, anger, sadness, fear and normal (unemotional) state. Five subjects recorded the sentences twice with different recording parameters. Thus, each subject recorded 20 or 40 utterances, yielding a corpus of 700 utterances[1], with 140 utterances per emotional state.

3.2 People Performance And Data Selection

We designed an experiment to answer the following questions: How well can people without special training portray and recognize emotions in speech? Which kinds of emotions are easier/harder to recognize?

We implemented an interactive program that selected and played back the utterances in random order and allowed a user to classify each utterance according to its emotional content. Twenty-three subjects took part in the evaluation stage, twenty of whom had participated in the recording stage earlier. Table 9.1 shows the performance confusion matrix[2]. We can see that the most easily recognizable category is anger (72.2%) and the least easily recognizable category is fear (49.5%). A lot of confusion is going on between sadness and fear, sadness and unemotional state, and happiness and fear. The mean accuracy is 63.5%, showing agreement with other experimental studies [11, 2].

Table 9.1. Performance Confusion Matrix.

Category	*Normal*	*Happy*	*Angry*	*Sad*	*Afraid*	*Total*
Normal	66.3	2.5	7.0	18.2	6.0	100%
Happy	11.9	61.4	10.1	4.1	12.5	100%
Angry	10.6	5.2	72.2	5.6	6.3	100%
Sad	11.8	1.0	4.7	68.3	14.3	100%
Afraid	11.8	9.4	5.1	24.2	49.5	100%

The left half of Table 9.2 shows statistics for evaluators for each emotion category. We can see that the variance for anger and sadness is significantly less than for the other emotion categories. This means that people better understand how to express/decode anger and sadness than other emotions. The right half of Table 9.2 shows statistics for "actors", i.e., how well subjects portray emotions. Comparing the left and right parts of Table 9.2, it is interesting to see that the ability to portray emotions (total mean is 62.9%) stays approximately at the same level as the ability to recognize emotions (total mean is 63.2%), but the variance for portraying is much larger.

From the corpus of 700 utterances we selected five nested data sets which include utterances that were recognized as portraying the given emotion by at least p per cent of the subjects (with p = 70, 80, 90, 95, and 100%). We will refer to these data sets as $s70$, $s80$, $s90$, $s95$, and $s100$. The sets contain

Table 9.2. Evaluators' and Actors' statistics.

	Evaluators' statistics					*Actors' statistics*				
Category	*Mean*	*s.d.*	*Median*	*Min*	*Max*	*Mean*	*s.d.*	*Median*	*Min*	*Max*
Normal	66.3	13.7	64.3	29.3	95.7	65.1	16.4	68.5	26.1	89.1
Happy	61.4	11.8	62.9	31.4	78.6	59.8	21.1	66.3	2.2	91.3
Angry	72.2	5.3	72.1	62.9	84.3	71.7	24.5	78.2	13.0	100
Sad	68.3	7.8	68.6	50.0	80.0	68.1	18.4	72.6	32.6	93.5
Afraid	49.5	13.3	51.4	22.1	68.6	49.7	18.6	48.9	17.4	88.0

the following number of items: $s70$: 369 utterances or 52.0% of the corpus; $s80$: 257/36.7%; $s90$: 149/21.3%; $s95$: 94/13.4%; and $s100$: 55/7.9%. We can see that only 7.9% of the utterances of the corpus were recognized by all subjects, and this number lineally increases up to 52.7% for the data set $s70$, which corresponds to the 70% level of concordance in decoding emotion in speech. Distribution of utterances among emotion categories for the data sets is close to a uniform distribution for $s70$ with $\sim$20% for normal state and happiness, $\sim$25% for anger and sadness, and 10% for fear. But for data sets with higher level of concordance anger begins to gradually dominate while the proportion of the normal state, happiness and sadness decreases. Interestingly, the proportion of fear stays approximately at the same level ($\sim$7–10%) for all data sets. The above analysis suggests that anger is easier to portray and recognize because it is easier to come to a consensus about what anger is.

3.3 Feature Extraction

All studies in the field point to pitch (fundamental frequency) as the main vocal cue for emotion recognition. Other acoustic variables contributing to vocal emotion signaling are [1]: vocal energy, frequency spectral features, formants (usually only one or two first formants (F1, F2) are considered), and temporal features (speech rate and pausing). Another approach to feature extraction is to enrich the set of features by considering some derivative features such as LPCC (linear predictive coding cepstrum) parameters of signal [12] or features of the smoothed pitch contour and its derivatives [5].

For our study we estimated the following acoustic variables: fundamental frequency F0, energy, speaking rate, and first three formants (F1, F2, and F3) and their bandwidths (BW1, BW2, and BW3), and calculated some descriptive statistics for them[3]. Then we ranked the statistics using feature selection techniques, and picked a set of most "important" features. We used the RELIEF-F algorithm [8] for feature selection[4] and identified 14 top features[5]. To investigate how sets of features influence the accuracy of emotion recognition algorithms we formed 3 nested sets of features based on their sum of ranks[6].

3.4 Computer Recognition

To recognize emotions in speech we tried the following approaches: K-nearest neighbors, neural networks, ensembles of neural network classifiers, and set of experts. In general, the approach that is based on ensembles of neural network recognizers outperformed the others, and it was chosen for implementation at the next stage. We summarize below the results obtained with the different techniques.

K-nearest neighbors. We used 70% of the $s70$ data set as database of cases for comparison and 30% as test set. We ran the algorithm for $K = 1$ to 15 and for number of features 8, 10, and 14. The best average accuracy of recognition (~55%) can be reached using 8 features, but the average accuracy for anger is much higher (~65%) for 10- and 14-feature sets. All recognizers performed very poor for fear (about 5–10%).

Neural networks. We used a two-layer backpropagation neural network architecture with a 8-, 10- or 14-element input vector, 10 or 20 nodes in the hidden sigmoid layer and five nodes in the output linear layer. To train and test our algorithms we used the data sets $s70$, $s80$ and $s90$, randomly split into training (70% of utterances) and test (30%) subsets. We created several neural network classifiers trained with different initial weight matrices. This approach applied to the $s70$ data set and the 8-feature set gave an average accuracy of about 65% with the following distribution for emotion categories: normal state is 55–65%, happiness is 60–70%, anger is 60–80%, sadness is 60–70%, and fear is 25–50%.

Ensembles of neural network classifiers. We used ensemble[7] sizes from 7 to 15 classifiers. Results for ensembles of 15 neural networks, the $s70$ data set, all three sets of features, and both neural network architectures (10 and 20 neurons in the hidden layer) were the following. The accuracy for happiness remained the same (~65%) for the different sets of features and architectures. The accuracy for fear was relatively low (35–53%). The accuracy for anger started at 73% for the 8-feature set and increased to 81% for the 14-feature set. The accuracy for sadness varied from 73% to 83% and achieved its maximum for the 10-feature set. The average total accuracy was about 70%.

Set of experts. This approach is based on the following idea. Instead of training a neural network to recognize all emotions, we can train a set of specialists or experts[8] that can recognize only one emotion and then combine their results to classify a given sample. The average accuracy of emotion recognition for this approach was about 70% except for fear, which was ~44% for the 10-neuron, and ~56% for the 20-neuron architecture. The accuracy of nonemotion (non-angry, non-happy, etc.) was 85–92%. The important question is how to combine opinions of the experts to obtain the class of a given sample. A simple and natural rule is to choose the class with the expert value closest to 1. This rule gives a total accuracy of about 60% for the 10-neuron architecture, and about 53% for the 20-neuron architecture. Another approach to rule selection is to use the outputs of expert recognizers as input vectors for a new neural network. In this case, we give the neural network the opportunity to learn itself the most appropriate rule. The total accuracy we obtained[9] was about 63% for both 10- and 20-node architectures. The average accuracy for sadness was rather high ($\sim$76%). Unfortunately, the accuracy of expert recognizers was not high enough to increase the overall accuracy of recognition.

4. Development

The following pieces of software were developed during the second stage: *ERG* – Emotion Recognition Game; *ER* – Emotion Recognition Software for call centers; and *SpeakSoftly* – a dialog emotion recognition program. The first program was mostly developed to demonstrate the results of the above research. The second software system is a full-fledged prototype of an industrial solution for computerized call centers. The third program just adds a different user interface to the core of the ER system. It was developed to demonstrate real-time emotion recognition. Due to space constraints, only the second software will be described here.

4.1 ER: Emotion Recognition Software For Call Centers

Goal. Our goal was to create an emotion recognition agent that can process telephone quality voice messages (8 kHz/8 bit) and can be used as a part of a decision support system for prioritizing voice messages and assigning a proper agent to respond the message.

Recognizer. It was not a surprise that anger was identified as the most important emotion for call centers. Taking into account the importance of anger and the scarcity of data for some other emotions, we decided to create a recognizer that can distinguish between two states: "agitation" which includes anger, happiness and fear, and "calm" which includes normal state and sadness. To create the recognizer we used a corpus of 56 telephone messages of varying length (from 15 to 90 seconds) expressing mostly normal and angry emotions that were recorded by eighteen non-professional actors. These utterances were automatically split into 1–3 second chunks, which were then evaluated and labeled by people. They were used for creating recognizers[10] using the methodology developed in the first study.

System Structure. The ER system is part of a new generation computerized call center that integrates databases, decision support systems, and different media such as voice messages, e-mail messages and a WWW server into one information space. The system consists of three processes: a wave file monitor, a voice mail center and a message prioritizer. The wave file monitor reads periodically the contents of the voice message directory, compares it to the list of processed messages, and, if a new message is detected, it processes the message and creates a summary and an emotion description file. The summary file contains the following information: five numbers that describe the distribution of emotions, and the length and percentage of silence in the message. The emotion description file stores data describing the emotional content of each 1–3 second chunk of message. The prioritizer is a process that reads summary files for processed messages, sorts them taking into account their emotional content, length and some other criteria, and suggests an assignment of agents to return back the calls. Finally, it generates a web page, which lists all current assignments. The voice mail center is an additional tool that helps operators and supervisors to visualize the emotional content of voice messages.

5. Conclusion

We have explored how well people and computers recognize emotions in speech. Several conclusions can be drawn from the above results. First, decoding emotions in speech is a complex process that is influenced by cultural, social, and intellectual characteristics of subjects. People are not perfect in decoding even such manifest emotions as anger and happiness. Second, anger is the most recognizable and easier to portray emotion. It is also the most important emotion for business. But anger has numerous variants (for example, hot anger, cold anger, etc.) that can bring variability into acoustic features and dramatically influence the accuracy of recognition. Third, pattern recognition techniques based on neural networks proved to be useful for emotion recognition in speech and for creating customer relationship management systems.

Notes

1. Each utterance was recorded using a close-talk microphone. The first 100 utterances were recorded at 22-kHz/8 bit and the remaining 600 utterances at 22-kHz/16 bit.

2. The rows and the columns represent true and evaluated categories, respectively. For example, the second row says that 11.9% of utterances that were portrayed as happy were evaluated as normal (unemotional), 61.4% as true happy, 10.1% as angry, 4.1% as sad, and 12.5% as afraid.

3. The speaking rate was calculated as the inverse of the average length of the voiced part of utterance. For all other parameters we calculated the following statistics: mean, standard deviation, minimum, maximum, and range. Additionally, for F0 the slope was calculated as a linear regression for voiced part of speech, i.e. the line that fits the pitch contour. We also calculated the relative voiced energy. Altogether we have estimated 43 features for each utterance.

4. We ran RELIEF-F for the $s70$ data set varying the number of nearest neighbors from 1 to 12, and ordered features according their sum of ranks.

5. The top 14 features are: F0 maximum, F0 standard deviation, F0 range, F0 mean, BW1 mean, BW2 mean, energy standard deviation, speaking rate, F0 slope, F1 maximum, energy maximum, energy range, F2 range, and F1 range.

6. The first set included the top 8 features (from F0 maximum to speaking rate), the second extended the first by the next 2 features (F0 slope and F1 maximum), and the third included all 14 top features.

7. An ensemble consists of an odd number of neural network classifiers trained on different subsets. The ensemble makes a decision based on the majority voting principle.

8. To train the experts, we used a two-layer backpropagation neural network architecture with a 8-element input vector, 10 or 20 nodes in the hidden sigmoid layer and one node in the output linear layer. We also used the same subsets of the $s70$ data set as training and test sets but with only two classes (for example, angry – non-angry).

9. To explore this approach, we used a two-layer backpropagation neural network architecture with a 5-element input vector, 10 or 20 nodes in the hidden sigmoid layer and five nodes in the output linear layer. We selected five of the best experts and generated several dozens neural network recognizers.

10. We created ensembles of 15 neural network recognizers for the 8-,10-, and 14-feature inputs and the 10- and 20-node architectures. The average accuracy of the ensembles of recognizers lies in the range 73–77% and achieves its maximum $\sim$77% for the 8-feature input and 10-node architecture.

References

[1] R. Banse and K.R. Scherer. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70: 614–636, 1996.

[2] R. van Bezooijen. *The characteristics and recognizability of vocal expression of emotions*. Foris, Drodrecht, The Netherlands, 1984.

[3] J.E. Cahn. Generation of Affect in Synthesized Speech. In *Proc. 1989 Conference of the American Voice I/O Society*, pages 251–256. Newport Beach, CA, September 11–13, 1989.

[4] C. Darwin. *The expression of the emotions in man and animals*. University of Chicago Press, 1965 (Original work published in 1872).

[5] F. Dellaert, T. Polzin, and A. Waibel. Recognizing emotions in speech. In *Proc. Intl. Conf. on Spoken Language Processing*, pages 734–737. Philadelphia, PA, October 3–6, 1996.

[6] C. Elliot and J. Brzezinski. Autonomous Agents as Synthetic Characters. *AI Magazine*, 19: 13–30, 1998.

[7] L. Hansen and P. Salomon. Neural Network Ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 12: 993–1001, 1990.

[8] I. Kononenko. Estimating attributes: Analysis and extension of RELIEF. In L. De Raedt and F. Bergadano, editors, *Proc. European Conf. On Machine Learning (ECML'94)*, pages 171–182. Catania, Italy, April 6–8, 1994.

[9] I.R. Murray and J.L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotions. *J. Acoust. Society of America*, 93(2): 1097–1108, 1993.

[10] R. Picard. *Affective computing*. MIT Press, Cambridge, MA, 1997.

[11] K.R. Scherer, R. Banse, H.G. Wallbott, and T. Goldbeck. Vocal clues in emotion encoding and decoding. *Motivation and Emotion*, 15: 123–148, 1991.

[12] N. Tosa and R. Nakatsu. Life-like communication agent: Emotion sensing character "MIC" and feeling session character "MUSE". In *Proc. Third IEEE Intl. Conf. on Multimedia Computing and Systems*, pages 12–19. Hiroshima, Japan, June 17–23, 1996.

Chapter 10

SOCIAL INTELLIGENCE FOR COMPUTERS

Making Artificial Entities Creative in their Interactions

Juliette Rouchier
GREQAM (CNRS)

Abstract I review two main principles that have been developed to coordinate artificial agents in Multi-Agent systems. The first is based on the elaboration of complex communications among highly cognitive agents. The other is eco-resolution, where very simple agents have no consciousness of the existence of others. Both approaches fail to produce a social life that is close to that of humans, in terms of creativity or exchange of abstractions. Humans can build new ways of communicating, even with unknown entities, because they suppose that the other is able to give a meaning to messages, and are able to transfer a protocol from one social field to another. Since we want social intelligence to be creative, it seems that a first step would be to have agents be willing to communicate and know more than one reason and way to do so.

1. Introduction

Here, I compare computers' social intelligence to the human one. There is no generally agreed definition of social intelligence, but several elements seem to be indicated by the difference between human intelligence and more basic cognitions. These include: the ability to communicate with others in order to undertake common actions; the ability displayed by a society to integrate newcomers (and conversely for individuals to adapt to new ways of interacting) in order to communicate with unknown people; the ability to understand what others want from you, how you can help, or conversely influence others so that they help you [1].

Some progress has recently been made towards the understanding of social intelligence for non-living entities in the field of Multi-Agent Systems (MAS). MAS focuses on the socialisation of artificial intelligences using a variety of approaches. Attempts to create a real artificial intelligence are often based upon the human perception of intelligence. The comparison of computers' sociality to the human one is sharp - one can see that humans are able to get involved in very complex behaviours when interacting, even when they have only a minimal amount of conscious data about the others.

If the final aim is to engender some social intelligence in artificial agents, the usual approaches (e.g. trying to limit misunderstandings in the interpretation of messages exchanged or by reducing uncertainty in social organisation) might not be the best way forward. These approaches do not contradict that aim, in fact they can be useful, but they will not be sufficient.

2. Socially Intelligence In MAS

The problem of organising interactions among computers has grown drastically since the spread of PCs to a wider public, and more precisely with the ever-wider adoption of computer networks. The field of Multi-Agent Systems that has developed in the last years was originally led by the idea of having several artificial intelligences accomplish tasks together. This is often seen as a continuation of both Artificial Intelligence and Distributed Artificial Intelligence, and researchers usually look for ways of building societies of artificial intelligences [12].

There exists a commonly accepted definition of a MAS: it is constituted of an environment that evolves in time in which autonomous entities (agents) are able to act; they can also interact; the overall organisation is effected by rules that coordinate time evolution and all actions of the entities. Four axes can thus be emphasised for such a systems, (called the AEIO model): the exploration of what an Agent should be (mainly in terms of cognition and perception), what characteristics the Environment has, which kind of Interactions can exist between the entities, and what kind of Organisation is needed to put these different elements together [2].

The reasons why people use MAS are diverse. Some need to coordinate computers, as in networks, some need to coherently integrate different expert systems, and some need to make robots act together. In most cases, it is common to consider that the reliability of the system, its predictability and its efficiency are what is most important. In order to coordinate agents so that they work together without interfering with each other it is common to use a set of norms. In that case agents have clear roles that are interdependent and a known set of acquaintances with complementary roles (for example [5]. In this case, communication is mainly used by agents to make requests for help and to accept tasks that are proposed to them. Since tasks are necessarily thought as collective, the social life of agents is very important for the global system. They are thus designed with this ability to interact, either by being committed to others

[20], or by having intentions towards the others [11]. In such an organisation, agents can choose whether or not to cooperate with the others but have no ability to make choices about their social life by, for example, choosing with whom they can interact, and on which topic. This limitation is often justified by the need for stability.

However, for some people it is important that relations can evolve over time, since not all agents in the network are reliable [18]. Some of these systems are designed so that the relations between agents evolve and agents build representations about their network. Whenever an agent engages in some work for or with another, both can change their point of view on the relation, and each one can decide whether to stop the interaction or reinforce some weaker link.

This form of learning could help in addressing the question of social order in a deeper way, but it is usually used in networks where the agents already know the whole group at the beginning or can get in touch with any other. Thus these systems don't display one of the main characteristics of social life, which is *openness*. This property is the ability to accept newcomers into a group, to have them be integrated into the usual activities and judged as the others are. This *openness* has often recognised as a very important question for MAS [4]. Some systems were hence designed with the aim of dealing with that openness. MadKit [13] is one of them: in this system agents are divided into communication groups. Most of the agents belong to just one group and are not necessarily aware that there exist other agents with whom they don't interact. Communication between groups is very important and is done by representative agents; these also receive requests from agents who are outside the group and ask for entry. If accepted, the requesting agent is allowed to be a normal agent among the others.

Although being quite open, the organisation is necessarily based on strong assumptions about the agents and their ability to interact. Agents still have predefined goals, the idea that they can ask for help is already implemented, and they express their needs to the others in a direct way (i.e they know with whom they are to communicate and how to interpret the messages from them). This implies at least a common language and similar cognitive processes. This reduces the flexibility of the evolving social life.

There is a different approach in MAS - one that draws its inspiration from societies of simple animals, rather than using linguistic metaphors of communication. This choice of bottom-up perspective is often referred to as "eco-resolution" and can exhibit quite complex social patterns that can be seriously validated [9]. It is possible to build systems where each agent has no global knowledge, communication between agents is not direct but takes place through the environment and the fulfilment of tasks mainly relies on self-organising properties of the system. Each agent acts with reflexes that are provoked by the discovery of stimuli in the environment, and the intensity and position of these stimuli evolves in relation to the resolution of some of the problems. As long as a task hasn't been fulfilled, it can still be perceived as a task to be done by an agent passing by; if another agent does it, the stimulus disappears or decreases. Agents learn, for example by the use re-enforcement according to tasks they have done already, and thus according to what was not done by other entities. The system is completely socially regulated, although not a single agent has the perception of the existence of other agents.

This is often considered as a radically opposed point of view to the preceding one. It has a great flexibility in terms of the openness in the system. On the other hand it is not easy to know how this approach can help to coordinate agents that need to exchange complex information or confidential information that cannot be abandoned in an open environment.

None of these approaches exhibits agents which combine the two different abilities: to be able to meet new entities with potentially different ways of communicating (openness) and integrate with them into a normal communication network so that they can exchange important information or consciously organise common work. This double social competence could be defined as the ability to build trust (a "coherent" trust: which doesn't affect the survival of the individual agents or the system) and is clearly hard to create with artificial entities.

3. Social Intelligence And Creativity

It could be argued that because the kind of inputs that humans get from their interactions are of diverse forms, are more complex and carry more information than written messages, that this explains why they are more able to make inferences about unknown people. In the description of human interactions, not only body movements and positions are studied but also geographical relative position and the use of time in relations - as can be seen with [15] and [16]. I claim that this argument is not relevant, since the characteristics of human interaction can be recognised in very artificial settings. Here the example of interactions among humans who use computer networks to communicate is relevant. A human looking at a screen has necessarily less data coming from the interaction channel than the computer itself, but he or she seems to be able do much more about it, and be able to turn this data into information. Two examples illustrate the social complexity that can emerge from exchanges over open networks: academic discussion lists, and communities of teenagers playing games on the Internet.

Academics frequently exchange points of view via the Internet and do so publicly in discussion lists. Watching the traffic on these lists, one can identify unofficial reasons that lead people to participate. These include: the creation of their own reputation and the discovery of allies. Sometimes, considering the violence of exchanges and the extreme positions participants take, one could say that the goal of exchanging ideas has been subsumed and remains merely to sustain the appearance of a meaningful social sub-system. Most of the time the type of interactions involved and the shape of the communication mirror classic power relations found in the "real" world.

When teenagers interact by participating to games over the Internet, they know very quickly who is reliable or not and sometimes, they choose to play a session just because they know one participant and want to play with him (rarely her). While playing, they can identify the type of behaviour the other person has and form a definite opinion of it, e.g. like it or dislike it. Even if they don't actually meet the people, they often mix relations and communicate more intimate information through ICQ, give advice to each other about how to find music, new free games, or videos. The relations are usually more flexible and less passionate than those of adults, maybe because younger people feel more comfortable with the computer tool itself, or because the relations they experience in real life are more relaxed as well.

The pattern of relations and the rules of communication develop in ways that are similar to that in the outside world, but not identical. They also constitute an independent history of computer network communication that is very creative in terms of language and social codes. Existing structures are important as they provide the foundations for interaction and a form of justification for it, but the actual establishment of alliances between individuals is mainly based on a new ability: the ability to use the norms of that special network to distinguish oneself or identify others.

Two elements that are at the basis of human socialisation can be recognised. An important ability of humans is the recognition of the other as a "self", similar to his own - this makes it possible to anticipate or interpret the other's actions by projecting one's own acts [19]. Secondly, humans can create new norms of communication in sub-groups, mainly by adapting norms that they already know: the mental environment that enabled them to learn how to communicate and be creative [21].

Identifying a person as being relevant for communication has not been successfully implemented in AI. It is closely related to the idea of creating an artificial consciousness - one that would be able to identify a meaningful event without having preconceptions on what to look for but only an ability to learn [6]. Since that kind of artificial system is not available yet, most of the work about "like-me" test [7] postulates the ability to recognise a relevant self. They develop the idea of misconceptions in projection that can lead to the creation of a strange social life, quite unpredictable, but capturing essential properties of human societies [8]. Others try to teach useless robots how to manipulate humans in order to keep themselves working: one can hope that by postulating this surprising egocentrism, the robots will build strong representations of their social situation, which enable them to act coherently at any moment, in the face of unexpected and complex situations [10].

The idea of creating new communication groups and norms from existing ones has been used for quite a long time: for example, by [17] and finding ways of having new institutions emerge is often emphasised (e.g. [14]. Although successful, these attempts still require very little creativity from the agents in their social life, since this ability would require the involvement of a meta-language.

It is obvious that the creativity that is required is not easy to organise. Just one element, inspired by the observation of human societies, could help in improving it. As we saw, humans use old rules to create new ways of communicating by putting them in a new context. It is true in everyday life: some power or trust relations that are expected to appear in intimate relations can sometimes be recognised at work. Even without much invention, this overlap between different fields of relation can eventually engender new institutions: the change of context forces individuals to explain to the others the meaning of the new rule they try to use and thus create a new relational system [3]. For the moment, software systems are built in a way that makes them very specialised and quite independent from each other. More generally, when agents communicate, it is mainly about single topics, rarely dealing about different type of tasks with different type of acquaintances. Maybe making agents more generic would be the first step to enable them to transpose their knowledge of interactions between different contexts so that the result would be more creative.

4. Conclusion

The idea of social intelligence can be summarised by two main requirements: the ability to be able to exchange quite complex information to undertake sophisticated tasks in a reliable way, but also the ability to open the system so that new rules of communication can be created with new type of agents when they are recognised as valuable for these interactions.

For the moment these aspects are treated quite independently, and the reason certainly lies in the final aims of the research. A technological system that has to be reliable cannot afford to address the questions of the self-consciousness of artificial agents: such approaches have not been proven to lead to predictable results (and one could even anticipate the opposite effect). This is why no one tries to design social intelligence of entities when these are used to solve precise problems.

Lots of researchers agree on the fact that social intelligence can appear only if agents can recognise who is valuable for interaction, and more precisely if they are willing to communicate with others even if they receive messages that are not clear right away. But in that case, one has to enable them to imagine new forms of communication, building ways to secure their trust in any newly met agent that must not stay a stranger.

I recalled that, in order to create new ways of communicating, even humans need to take inspiration in existing institutions in order to create new relational patterns. What is important in this case is the transposition of these institutions in a new field of relation. It thus seems reasonable to argue that, in the case artificial agents, transposition of old communication systems (that don't need to be non-contradictory) in a new context could also be at the basis of the creativity we are looking for. The actual research on agents languages, trying to reduce ambiguity in communication, may at some point help to design socially intelligent agents by giving them examples of what communication is, before they produce alternative ways. But at the same time it stays clear that specialisation in a task is contradictory to the presence of creativity in social relations. The desire for communication, a range of diverse example of quite sophisticated interactions and a huge number of reasons to communicate among themselves, seem to be necessary to sustain artificial agents in their attempt to find out the intention of the others and adapt to their habits of communication constantly.

References

[1] Byrne R., Whiten A. *Machiavellian Intelligence.* Clarendon Press, Oxford, 1988.

[2] Boissier O., Demazeau Y., Sichman J. Le problème du contrôle dans un Système Multi-Agent (vers un modèle de contrôle Social). In: 1ère Journée Nationale du PRC-IA sur les Systèmes Multi-Agents, Nancy, 1992.

[3] Boltanski L., Thévenot L. *De la justification : les économies de la grandeur*, Gallimard, Paris, 1987.

[4] Bordini R. H. Contributions to an Anthropological Approach to the Cultural Adaptation of Migrant Agents, University College, London, Department of Computer science, University of London, 1999.

[5] Castelfranchi C., Conte R. Distributed Artificial Intelligence and social science: critical issues, In: *Foundations in Distributed Artificial Intelligence*, Wiley, 527–542, 1996.

[6] Cardon A. Conscience artificielle et systèmes adaptatifs. Eyrolles, Paris, 1999.

[7] Dautenhahn, K. I Could Be You: The phenomenological dimension of social understanding. *Cybernetics and Systems*, 28:417–453, 1997.

[8] Doran J. Simulating Collective Misbelief, *Journal of Artificial Societies and Social Simulation*, <http://www.soc.surrey.ac.uk/JASSS/1/1/3.html>, 1998.

[9] Drogoul A., Corbara B., Lalande S. MANTA: new experimental results on the emergence of (artificial) ant societies, In: *Artificial societies. The computer simulation of social life*, UCL, London, pp 190–211, 1995.

[10] Drogoul A. Systèmes multi-agents situés. Mémoire d'habilitation à diriger des recherches (habilitation thesis), 17 march 2000.

[11] Esfandiari B., et al.. Systèmes Multi-Agents et gestion de réseaux, Actes des 5ème Journées Internationales PRC-GDR Intelligence Artificielle, , Toulouse, 317–345, 1995.

[12] Ferber J. *Multi-Agent System. An introduction to Distributed Artificial Intelligence*, Addison-Wesley, England, 1999.

[13] Ferber J., Gutknecht O. A Meta-Model for the Analysis and Design of Organizations in Multi-Agent Systems, Proc. ICMAS '98, IEEE Comp. Soc., 128–135, 1998.

[14] Glance N.S, Huberman B.A. Organizational fluidity and sustainable cooperation. In: *From Reaction to Cognition*, Castelfranchi C., Muller J.P. (eds) LNAI, 957, Springer, Heidelberg, 89–103, 1995.

[15] Goffman E. *Encounters. Two Studies in the Sociology of interaction*, Bobbs-Herrill Company, Indianapolis, 1965.

[16] Hall E. T. *The hidden dimension: man's use of space in public and private*, Bodley Head, London, 1969.

[17] Hirayama K., Toyoda J. Forming coalition for breaking deadlocks In: Proc. of ICMAS'95, MIT Press, Cambridge, 155–160, 1995.

[18] Marsh S., A Community of Autonomous Agents for the Search and Distribution of Information in Networks, 19th BCS-IRSG, 1997.

[19] Premack D. Does the Chimpanzee Have a Theory of Mind ?' Revisited, In: Richard Byrne and Andrew Whiten (eds) *Machiavellian Intelligence. Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, Clarendon Press, Oxford, pp 160–179, 1988.

[20] Sichman J., et al. Reasoning about others using dependence networks, 3rd Italian Workshop on D.A.I., AI*IA, Rome, 1993.

[21] Vygotsky L.S. Mind in Society, In: Michael Coke et al. (eds.), Harvard University Press, Cambridge, Massachusetts, 1978.

Chapter 11

EGOCHAT AGENT

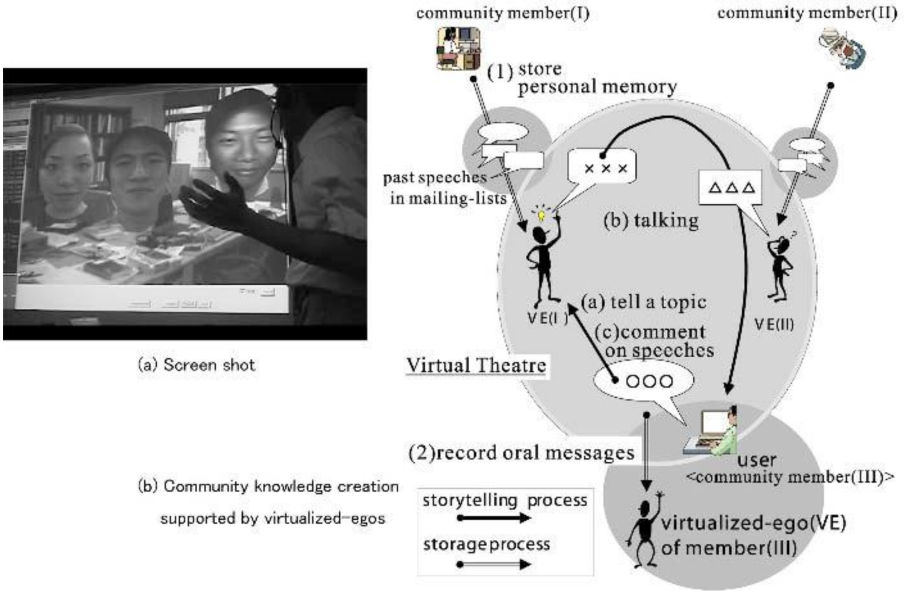
A Talking Virtualized Agent that Supports Community Knowledge Creation

Hidekazu Kubota and Toyoaki Nishida
The University of Tokyo, Faculty of Engineering

Abstract This paper proposes a method that supports knowledge creation in a community by mean of talking virtualized-egos. A virtualized-ego is a conversational agent that represents a community member. Community members can exchange their knowledge by talking with virtualized-egos even when some members are away because the virtualized-egos replay appropriate past messages posted by the absent members. Moreover, messages of virtualized-egos are ordered as a story. A story is a series of comments that reflects relationships between community members. We have developed a system called EgoChat to investigate these ideas and carried out an experiment to ascertain its effectiveness for community knowledge creation.

1. Introduction

In the human community, intelligent agents can facilitate community activities by interacting with humans. Such social agents have been studied in the research area called Socially Intelligent Agents (SIA) [2]. This paper presents a social agent that tells autobiographical stories on behalf of a community member and supports community knowledge creation. Among the human processes involved in knowledge creation, informal communication aimed at relating personal experiences results in the creation of innovative knowledge [7]. In the same way, interaction among humans and agents can also be enriched by personal experiences. Telling autobiographical stories about oneself promotes social understanding between humans and agents, thus improving relationships between them [3]. The great challenge in SIA research that we address in this paper is how to generate autobiographical conversations from our daily e-mails. Because the recent progress in communication technologies on electronic media such as e-mail or the WWW has made participating in a community much


(a) Screen shot

(b) Community knowledge creation supported by virtualized-egos

Figure 11.1. Overview of EgoChat system

easier, we apply SIAs on these electronic communication media. Our method for supporting knowledge creation is based on human-style talking agents called “virtualized-egos” (VEs). A VE is one’s other self; it works independently of one’s real self and can talk on one’s behalf. A VE compiles everyday e-mails as an autobiographical memory that stores experiences of a community member and generates a story from this memory. VEs work in our system called EgoChat, which is a virtual environment for a conversation among humans and VEs with voices. With EgoChat, community members can exchange their experiences with each other even if some members are absent because VEs can talk on behalf of the absent members.

2. Overview of EgoChat

Figure 11.1(a) shows a screen shot of EgoChat. A user is talking with three VEs, each of which has the 3-D head of another community member.

Speeches of VEs are ordered as a story. In this research, we define a story as a series of comments that represents relationships between community members. When we examine conversations in a mailing-list, we find that a members having the same interests are inclined to talk with one another. We therefore assumed that members in a general community have the same inclination, and that this inclination represents the relationships of community members.

The EgoChat system consists of three storytelling processes and two storage processes (Figure 11.1(b)). In the storytelling processes, community members can share their knowledge by creating a knowledge stream with their VEs. When a user wants to some community knowledge, she inputs her interest into EgoChat by voice [storytelling process (a)]. The voice messages are recognized by a commercial speech recognition system, and then VEs start talking about related topics [storytelling process (b)]. VEs will continue talking with one another unless the user interrupts the conversation to make a comment or change the subject [storytelling process (c)].

As the community member talks with VEs, her personal memory is enriched in the processes of storing comments. Before using EgoChat, the personal memories are stored in VEs by using automated summarizing technology or summarizing humanly from past exchanges on electronic media such as mailing-lists [storage process (1)]. Besides text-based messages, VEs store new oral messages from a user on EgoChat [storage process (2)]. In this way, VEs share community knowledge in past messages of community members and users add new ideas to their virtualized-egos in a loop of community knowledge creation.

3. Storytelling by VEs

VEs generate messages in turn, and these messages are ordered as a story as follows.

1. Start a turn. taking turns
2. Each VE selects a message associated with a topic from personal memory.
3. Only a VE whose message is presented in the most orderly and most reasonable way in the context speaks the message.
4. The turn ends, and the next begins.

By repeating the above process of generating and selecting messages, VEs can tell a story.

In the following three sections, we propose two sets of representations of the personal memory that characterize a personality of a VE as an agent of a community member; and lay out how an appropriate VE is selected during process 3.

3.1 Topics-and-summaries representation

Each VE has a set of topics. Past posts from a community member related to each topic are filed away into the personal memory of a VE. For example, when a VE in a community of liquor fans has topics such as brandy, beer and sake (Figure 11.2(a)), comments about brandy like “I used to drink V.S.O.P”

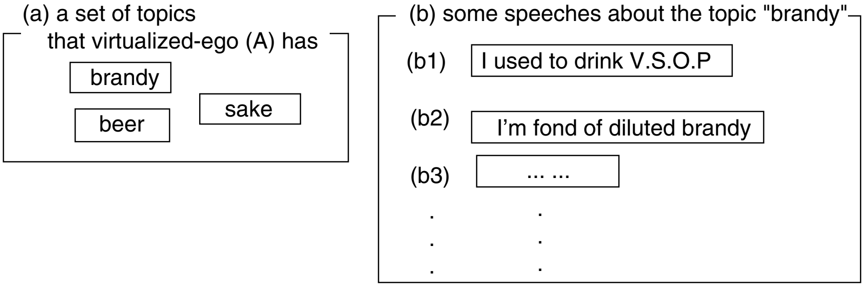

(a) a set of topics
that virtualized-ego (A) has
brandy
sake
beer
(b) some speeches about the topic "brandy"
(b1) I used to drink V.S.O.P
(b2) I'm fond of diluted brandy
(b3) ... ...

Figure 11.2. Example of a topics-and-summaries set

or "I'm fond of diluted brandy (Figure 11.2(b)) are stored in the memory of the VE and posts about other topics are stored in the same way.

3.2 Flow-of-topics representation

VEs change topics occasionally by referring to an associative representation set of a flow of topics. The associative representation proposed for the CoMeMo-Community [5] consists of many-to-many hyperlinks that associate one or more key unit with one or more value unit. The semantics of the associative representation are not defined strictly. Instead, we leave the interpretation of the semantics to human association based on our tacit background knowledge.

In the case of VE (a) (Figure 11.3), 'liquor' is a key unit and 'brandy' and 'beer' are value units. This associative representation of VE (a) shows a flow of topics from liquor to brandy or beer, and other associative representations such as that for VE (b), shows other flows of topics.

Associative representations that show associations of a community member are stored with "topics and summaries" in the memory of VE. One mediator selected by a user among VEs selects the next topic that is associated with the current topic. For instance, the message for changing topics could be; "I associate liquor with brandy. Next, let's talk about brandy." We believe that association-based flows of topics make the storytelling of VEs human-like and help users to view VEs as the independently working other selves of community members.

3.3 Storytelling by ordered messages

In a turn, though all the VEs select messages associated with a topic at the same time, only one VE is selected to speak at a time, which is done by comparing the priorities of selected messages. Each VE generates a priority when it selects a message. The criteria to decide priorities of VEs are as follows:

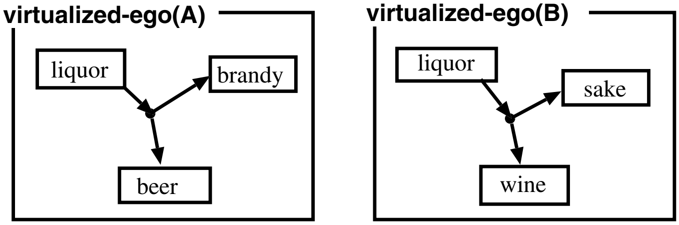


Figure 11.3. Example of a flow-of-topics set

Story structure: A stream of messages reflects social relationships between community members. The frequency of exchange between community members on mailing-lists and between a user and a VE on EgoChat is recorded in each VE. For example, when community member (a) is inclined to talk with member (b) more than with community member (c) and the VE of member (b) has talked in the previous turn, the VE of member (a), not that of member (c), goes first in the present turn.

Coherence: A VE that selects the following message after a message mentioned just before goes first so that messages are exchanged coherently.

The summaries in a personal memory are labeled with key words that represent the contents of the summaries. A message is regarded as the one that follows the previous message when its keyword matches the previous one.

Fairness: A VE that speaks little goes before one that speaks a lot for the sake of fairness.

4. Experiment

We carried out a basic experiment to ascertain the usability of the EgoChat system and investigate the effects of voice interaction between humans and agents. The experimental system was implemented in Java and Java3D API on a MS-Windows OS. The voices of VEs were not generated by text-to-speech software. Instead, they were recorded human voices. The body of a VE consists of a spherical head and a cone-shaped torso, and the head nods while talking.

4.1 Method

We created four VEs and generated their personal memories from a humanly summarized log of a mailing list about liquor where participants exchanged ideas about how to market brandy. Each VE represents a mailing-list participant. The subjects were three postgraduate students in our laboratory who have ordinary knowledge about liquor, and they are not members of the mailing list. Each subject was shown the VEs' conversation about brandy for three minutes and could comment on a point made by a VE when s/he feels an interest in the speech anytime.

The following is part of a conversation between VEs A, B, C and D and a subject[1].

A: "Let's talk about brandy."
B: "We have more prejudice against brandy than other liquor."
A: "I often chat over drinking."
C: "A lot of bars don't serve brandy because it is too expensive."
B: "Yes, certainly."
A: "Sure."
Subject: "Convenience stores don't deal in brandy either."
C: "Shouchu[2] is cheaper than brandy."

In the conversation, line 4 follows line 3 mainly because of the coherence criterion for storytelling. Both messages have been humanly labeled with the same key word "shouchu" in advance since these two messages originally appeared in the context of the advantages of shouchu. And the short responses "Yes, certainly" and "Sure" are randomly inserted in place of other messages from the personal memories to make the conversation rhythmic.

After the conversation, the subjects were asked whether they thought it was easier to comment on the messages of other persons using text-based mailing lists or the EgoChat system.

4.2 Results and Discussion

Two subjects answered that the EgoChat system was easier and more casual than a mailing list because a chat-style conversation with voices facilitated their interaction with VEs. This suggests that using storytelling agents with voices may facilitate interaction between humans and agents. For our previous system, CoMeMo-Community, we evaluated to what extent people shared their knowledge by the virtual exchange of messages by text-based words and images. The results for CoMeMo-Community were satisfactory [4]. Hence, we expect that the EgoChat, owing to its use of more human-like modalities such as voice and gestures, will bring community members more knowledge than CoMeMo-Community.

One subject answered that interacting with VEs is too unnatural for smooth communication because VEs can't answer to his questions. Certainly, the storytelling method on EgoChat is not powerful enough to give concise responses to users. On the other hand, EgoChat can easily store and process a great number of messages because the representation of personal memories and the methods of ordering messages are very simple. We are planning to practically apply EgoChat to a large community over a network and evaluate the amount of knowledge created in the community. In such an application, EgoChat has the merit of being able to deal with large amounts of data.

We foresee a lot of potential application areas, including campus communities made up of professors and students, knowledge management in a company, local commercial communities, and communities of volunteers. Currently, the EgoChat system is used in a public opinion channel (POC)[6], which is a novel communication medium for sharing and exchanging opinions in a community. POC is a kind of broadcasting system for a community that collects messages from members and feeds edited stories back to them. EgoChat plays mainly a broadcasting role in POC.

5. Related Work

Technologies for conversational agents that support our thinking process have been discussed in some works. The SAGE [8] agent helps to generate an autobiography of a user to enable the user to explore his inner world. Rea [1] and Imp Characters[3] work with human-like interaction to explain commercial information to customers. Each agent in these studies works alone and talks from one point of view. In contrast, EgoChat works with many agents and the agents talk from various view points of the community members. Therefore, the user can get many views about a topic on EgoChat.

6. Conclusion

We proposed a new method for supporting community knowledge creation. We have developed a system called EgoChat, which is a new asynchronous communication channel for a community that has a casual and conversational flavor. The virtual chat that the EgoChat system provides is applicable to a community formed over an asynchronous media, i.e., mailing lists or electronic bulletin boards. As future work, we are planning to apply EgoChat to a large community over a network and evaluate its usefulness.

Notes

1. The conversation was originally in Japanese.
2. Japanese distilled liquor made from wheat or rice.
3. Extempo, http://www.extempo.com/

References

[1] J. Cassell, H. Vilhjlmsson, K. Chang, T. Bickmore, L. Campbell, and H. Yan. Requirements for an architecture for embodied conversational characters. In *Computer Animation and Simulation '99 (Eurographics Series)*, 1999.

[2] Kerstin Dautenhahn. The art of designing socially intelligent agents - science, fiction, and the human in the loop. *Special Issue "Socially Intelligent Agents", Applied Artificial Intelligence Journal*, 12(7-8):573 – 617, 1998.

[3] Kerstin Dautenhahn. Story-telling in virtual environments. In *Working Notes Intelligent Virtual Environments, Workshop at the 13th biennial European Conference on Artificial Intelligence (ECAI-98)*, 1998.

[4] Takashi Hirata and Toyoaki Nishida. Supporting community knowledge evolution by talking-alter-egos metaphor. In *The 1999 IEEE Systems, Man, and Cybernetics Conference (SMC'99)*, 1999.

[5] Toyoaki Nishida. Facilitating community knowledge evolution by talking vitrualized egos. In Hans-Joerg Bullinger and Juegen Ziegler, editors, *Human-Computer Interaction VOLUME 2*, pages 437–441. Lawrence Erlbaum Associates, Pub., 1999.

[6] Toyoaki Nishida, Nobuhiko Fujihara, Shintaro Azechi, Kaoru Sumi, and Hiroyuki Yano. Public opinion channel for communities in the information age. *New Generation Computing*, 14(4):417–427, 1999.

[7] Ikujiro Nonaka and Hirotaka Takeuchi. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation.* Oxford University Press, 1995.

[8] Marina Umaschi. Sage storytellers: Learning about identity, language and technology. In *Proceedings of ICLS '96, AACE*, pages 526 – 531, 1996.

Chapter 12

ELECTRIC ELVES

Adjustable Autonomy in Real-World Multi-Agent Environments

David V. Pynadath and Milind Tambe
University of Southern California Information Sciences Institute

Abstract Through *adjustable autonomy* (AA), an agent can dynamically vary the degree to which it acts autonomously, allowing it to exploit human abilities to improve its performance, but without becoming overly dependent and intrusive. AA research is critical for successful deployment of agents to support important human activities. While most previous work has focused on individual agent-human interactions, this paper focuses on *teams* of agents operating in real-world human organizations, as well as the novel AA coordination challenge that arises when one agent's inaction while waiting for a human response can lead to potential miscoordination. Our multi-agent AA framework, based on Markov decision processes, provides an adaptive model of users that reasons about the uncertainty, costs, and constraints of decisions. Our approach to AA has proven essential to the success of our deployed Electric Elves system that assists our research group in rescheduling meetings, choosing presenters, tracking people's locations, and ordering meals.

1. Introduction

Software agents support critical human activities in intelligent homes [6], electronic commerce [2], long-term space missions [3], etc. Future human organizations will be even more highly agentized, with software agents supporting information gathering, planning, and execution monitoring, as well as having increased control of resources and devices. This agentization will assist organizations of all types, whether military, corporate, or educational. For example, in a research institution, agentization may facilitate meeting organization, paper composition, software development, etc. We envision agent proxies for each person within an organization. Thus, for instance, if an organization requires a deployment of people and equipment, then agent proxies could volunteer on behalf of the people or resources they represent, while also ensuring that the selected team collectively possesses sufficient resources and capabilities. The proxies could also monitor the progress of the participants and of the mission as a whole, executing corrective actions when necessary.

Applications of agents within human organizations have fostered an increasing interest in *adjustable autonomy* (AA), where agents dynamically adjust their own level of autonomy, harnessing human skills and knowledge as appropriate, without overly burdening the humans. When agents are embedded in large human organizations, they must also coordinate with each other and act jointly in teams. The requirements of teamwork and coordination give rise to novel AA challenges not addressed by previous research, which focuses on interactions between only an individual agent and its human user [3, 4, 5]. In particular, the *AA coordination challenge* arises during the transfer of decision-making control. In a team setting, an agent cannot transfer control freely, because as the agent waits for a human response, its teammates expect it to still fulfill its responsibilities to the overall joint task. Thus, the AA coordination challenge requires that an agent weigh possible team miscoordination while waiting for a human response against possible erroneous actions as a result of uninformed decisions.

We have conducted our research on AA using a real-world multi-agent system, *Electric Elves* (E-Elves) [1], that we have used since June 1, 2000, at USC/ISI. E-Elves assists a group of 10 users in their daily activities. To address the AA coordination challenge, E-Elves agents use Markov decision processes (MDPs) [8] to explicitly reason about team coordination via a novel three-step approach. First, before transferring decision-making control, an agent explicitly weighs the cost of waiting for user input and any potential team miscoordination against the cost of erroneous autonomous action. Second, agents do not rigidly commit to transfer-of-control decisions (as is commonly done in previous work), but instead reevaluate decisions as required. Third, an agent can change coordination arrangements, postponing or reordering activities, to "buy time" to lower decision cost/uncertainty. Overall, the agents look ahead at possible sequences of coordination changes, selecting one that maximizes team benefits.

2. Electric Elves

As a step towards agentization of large-scale human organizations, the Electric Elves effort at USC/ISI has had an agent team of 15 agents, including 10 proxies (for 10 people), running 24/7 since June 1, 2000, at USC/ISI [1]. The 5 other agents provide additional functionality for matching users' interests and capabilities and for extracting information from Web sites. Each agent proxy is called Friday (from Robinson Crusoe's servant Friday) and acts on behalf of its user in the agent team. If a user is delayed to a meeting, Friday can reschedule the meeting, informing other Fridays, who in turn inform their human users. If there is a research presentation slot open, Friday may respond to the invitation to present on behalf of its user. Friday can also order its user's meals and track the user's location. Friday communicates with users using wireless devices, such as Palm Pilots and WAP-enabled mobile phones, and via user workstations. We have used Friday's location reasoning to construct a People Locator that publishes the whereabouts of members of our research group on a Web page. This automatically updated information provides a cheap means for increasing social awareness (similar to previous work in the field [12]).

AA is of critical importance in Friday agents. Clearly, the more autonomous Friday is, the more time it saves its user. However, Friday has the potential to make costly mistakes when acting autonomously (e.g., volunteering an unwilling user for a presentation). Thus, each Friday must make intelligent decisions about when to consult its user and when to act autonomously. Furthermore, Friday faces significant, unavoidable uncertainty (e.g., if a user is not at the meeting location at meeting time, does s/he plan to attend?).

In addition to uncertainty and cost, the E-Elves domain raises the AA coordination challenge. Suppose that, when faced with uncertainty, a Friday agent consults its user (e.g., to check whether the user plans to attend a meeting), but the user, caught in traffic, fails to respond. While waiting for a response, Friday may miscoordinate with its teammates (other Friday agents), since it fails to inform them whether the user will attend the meeting. This, in turn means that other meeting attendees (humans) waste their time waiting. Conversely, if, to maintain coordination, Friday tells the other Fridays that its user will not attend the meeting, but the user does indeed plan to attend, the human team suffers a potentially serious cost from receiving this incorrect information. Friday must instead make a decision that makes the best tradeoff possible between the possible costs of inaction and the possible costs of incorrect action.

3. Decision-Tree Approach to AA

Our first attempt at AA in E-Elves was inspired by CAP [7], an agent system for helping a user schedule meetings. Like CAP, Friday learned user preferences using C4.5 decision-tree learning [9]. Although initial tests were promising [11], when we deployed the resulting system 24/7, it led to some dramatic failures, including:

1 Tambe's Friday incorrectly, autonomously cancelled a meeting with the division director. C4.5 over-generalized from training examples.

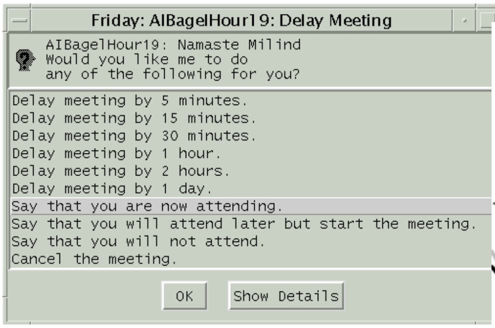
2 Pynadath's Friday incorrectly cancelled a meeting. A time-out forced the choice of an (incorrect) autonomous action when Pynadath did not respond.

3 A Friday delayed a meeting almost 50 times, each time by 5 minutes, ignoring the nuisance to the rest of the meeting participants.

4 Tambe's proxy automatically volunteered him for a presentation, though he was actually unwilling. Again, C4.5 had over-generalized from a few examples and when a timeout occurred had taken an undesirable autonomous action.

From the growing list of failures, it became clear that the approach faced some fundamental problems. The first problem was the AA coordination challenge. Learning from user input, when combined with timeouts, failed to address the challenge, since the agent sometimes had to take autonomous actions although it was ill-prepared to do so (examples 2 and 4). Second, the approach did not consider the team cost of erroneous autonomous actions (examples 1 and 2). Effective agent AA needs explicit reasoning and careful tradeoffs when dealing with the different individual and team costs and uncertainties. Third, decision-tree learning lacked the lookahead ability to plan actions that may work better over the longer term. For instance, in example 3, each five-minute delay is appropriate *in isolation*, but the rules did not consider the ramifications of one action on successive actions. Planning could have resulted in a one-hour delay instead of many five-minute delays. Planning and consideration of cost could also lead to an agent taking the low-cost action of a short meeting delay while it consults the user regarding the higher-cost cancel action (example 1).

4. MDPs for Adjustable Autonomy



Friday: AIBagelHour19: Delay Meeting

```
AIBagelHour19: Namaste Milind
Would you like me to do
any of the following for you?
```

```
Delay meeting by 5 minutes.
Delay meeting by 15 minutes.
Delay meeting by 30 minutes.
Delay meeting by 1 hour.
Delay meeting by 2 hours.
Delay meeting by 1 day.
Say that you are now attending.
Say that you will attend later but start the meeting.
Say that you will not attend.
Cancel the meeting.
```

Figure 12.1. Dialog for meetings

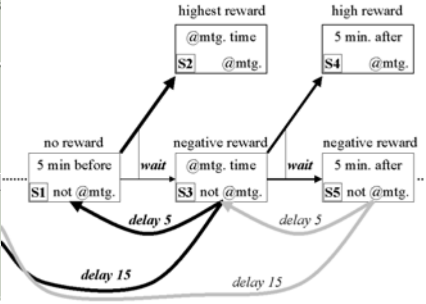


Figure 12.2. A small portion of simplified version of the delay MDP

MDPs were a natural choice for addressing the issues identified in the previous section: reasoning about the costs of actions, handling uncertainty, planning for future outcomes, and encoding domain knowledge. The delay MDP, typical of MDPs in Friday, represents a class of MDPs covering all types of meetings for which the agent may take rescheduling actions. For each meeting, an agent can autonomously perform any of the 10 actions shown in the dialog of Figure 12.1. It can also wait, i.e., sit idly without doing anything, or can reduce its autonomy and ask its user for input.

The delay MDP reasoning is based on a world state representation, the most salient features of which are the user's location and the time. Figure 12.2 shows a portion of the state space, showing only the location and time features, as well as some of the state transitions (a transition labeled "delay n" corresponds to the action "delay by n minutes"). Each state also has a feature representing the number of previous times the meeting has been delayed and a feature capturing what the agent has told the other Fridays about the user's attendance. There are a total of 768 possible states for each individual meeting.

The delay MDP's reward function has a maximum in the state where the user is at the meeting location when the meeting starts, giving the agent incentive to delay meetings when its user's late arrival is possible. However, the agent could choose arbitrarily large delays, virtually ensuring the user is at the meeting when it starts, but forcing other attendees to rearrange their schedules. This team cost is considered by incorporating a negative reward, with magnitude proportional to the number of delays so far and the number of attendees, into the delay reward function. However, explicitly delaying a meeting may benefit the team, since without a delay, the other attendees may waste time waiting for the agent's user to arrive. Therefore, the delay MDP's reward function includes a component that is negative in states after the start of the meeting if the user is absent, but positive otherwise. The reward function includes other components as well and is described in more detail elsewhere [10].

The delay MDP's state transitions are associated with the probability that a given user movement (e.g., from office to meeting location) will occur in a given time interval. Figure 12.2 shows multiple transitions due to a 'wait' action, with the relative thickness of the arrows reflecting their relative probability. The "ask" action, through which the agent gives up autonomy and queries the user, has two possible outcomes. First, the user may not respond at all, in which case, the agent is performing the equivalent of a "wait" action. Second, the user may respond, with one of the 10 responses from Figure 12.1. A communication model [11] provides the probability of receiving a user's response in a given time step. The cost of the "ask" action is derived from the cost of interrupting the user (e.g., a dialog box on the user's workstation is cheaper than sending a page to the user's cellular phone). We compute the expected value of user input by summing over the value of each possible response, weighted by its likelihood.

Given the states, actions, probabilities, and rewards of the MDP, Friday uses the standard value iteration algorithm to compute an optimal policy, specifying, for each and every state, the action that maximizes the agent's expected utility [8]. One possible policy, generated for a subclass of possible meetings, specifies "ask" and then "wait" in state **S1** of Figure 12.2, i.e., the agent gives up some autonomy. If the world reaches state **S3**, the policy again specifies "wait", so the agent continues acting without autonomy. However, if the agent then reaches state **S5**, the policy chooses "delay 15", which the agent then executes autonomously. However, the exact policy generated by the MDP will depend on the exact probabilities and costs used. The delay MDP thus achieves the first step of Section 1's three-step approach to the AA coordination challenge: balancing individual and team rewards, costs, etc.

The second step of our approach requires that agents avoid rigidly committing to transfer-of-control decisions, possibly changing its previous autonomy decisions. The MDP representation supports this by generating an autonomy *policy* rather than an autonomy *decision*. The policy specifies optimal actions for each state, so the agent can respond to any state changes by following the policy's specified action for the new state (as illustrated by the agent's retaking autonomy in state **S5** by the policy discussed in the previous section). In this respect, the agent's AA is an ongoing process, as the agent acts according to a policy throughout the entire sequence of states it finds itself in.

The third step of our approach arises because an agent may need to act autonomously to avoid miscoordination, yet it may face significant uncertainty and risk when doing so. In such cases, an agent can carefully plan a change in coordination (e.g., delaying actions in the meeting scenario) by looking ahead at the future costs of team miscoordination and those of erroneous actions. The delay MDP is especially suitable for producing such a plan because it generates policies after looking ahead at the potential outcomes. For instance, the delay MDP supports reasoning that a short delay buys time for a user to respond, reducing the uncertainty surrounding a costly decision, albeit at a small cost.

Furthermore, the lookahead in MDPs can find effective long-term solutions. As already mentioned, the cost of rescheduling increases as more and more such repair actions occur. Thus, even if the user is very likely to arrive at the meeting in the next 5 minutes, the uncertainty associated with that particular state transition may be sufficient, when coupled with the cost of subsequent delays if the user does not arrive, for the delay MDP policy to specify an initial 15-minute delay (rather than risk three 5-minute delays).

5. Evaluation of Electric Elves

We have used the E-Elves system within our research group at USC/ISI, 24 hours/day, 7 days/week, since June 1, 2000 (occasionally interrupted for bug fixes and enhancements). The fact that E-Elves users were (and still are) willing to use the system over such a long period and in a capacity so critical to their daily lives is a testament to its effectiveness. Our MDP-based approach to AA has provided much value to the E-Elves users, as attested to by the 689 meetings that the agent proxies have monitored over the first six months of execution. In 213 of those meetings, an autonomous rescheduling occurred, indicating a substantial savings of user effort. Equally importantly, humans are also often intervening, leading to 152 cases of user-prompted rescheduling, indicating the critical importance of AA in Friday agents.

The general effectiveness of E-Elves is shown by several observations. Since the E-Elves deployment, the group members have exchanged very few email messages to announce meeting delays. Instead, Fridays autonomously inform users of delays, thus reducing the overhead of waiting for delayed members. Second, the overhead of sending emails to recruit and announce a presenter for research meetings is now assumed by agent-run auctions. Third, the People Locator is commonly used to avoid the overhead of trying to manually track users down. Fourth, mobile devices keep us informed remotely of changes in our schedules, while also enabling us to remotely delay meetings, volunteer for presentations, order meals, etc. We have begun relying on Friday so heavily to order lunch that one local Subway restaurant owner even suggested marketing to agents: *"More and more computers are getting to order food, so we might have to think about marketing to them!!"*

Most importantly, over the entire span of the E-Elves' operation, the agents have *never* repeated any of the catastrophic mistakes that Section 3 enumerated in its discussion of our preliminary decision-tree implementation. For instance, the agents do not commit error 4 from Section 3 because of the domain knowledge encoded in the bid-for-role MDP that specifies a very high cost for erroneously volunteering the user for a presentation. Likewise, the agents never committed errors 1 or 2. The policy described in Section 4 illustrates how the agents would first ask the user and then try delaying the meeting, before taking any final cancellation actions. The MDP's lookahead capability also prevents the agents from committing error 3, since they can see that making one large delay is preferable, in the long run, to potentially executing several small delays. Although the current agents do occasionally make mistakes, these errors are typically on the order of asking the user for input a few minutes earlier than may be necessary, etc. Thus, the agents' decisions have been reasonable, though not always optimal. Unfortunately, the inherent subjectivity in user feedback makes a determination of optimality difficult.

6. Conclusion

Gaining a fundamental understanding of AA is critical if we are to deploy multi-agent systems in support of critical human activities in real-world settings. Indeed, living and working with the E-Elves has convinced us that AA is a critical part of any human collaboration software. Because of the negative result from our initial C4.5-based approach, we realized that such real-world, multi-agent environments as E-Elves introduce novel challenges in AA that previous work has not addressed. For resolving the *AA coordination challenge*, our E-Elves agents explicitly reason about the costs of team miscoordination, they flexibly transfer autonomy rather than rigidly committing to initial decisions, and they may change the coordination rather than taking risky actions in uncertain states. We have implemented our ideas in the E-Elves system using MDPs, and our AA implementation nows plays a central role in the successful 24/7 deployment of E-Elves in our group. Its success in the diverse tasks of that domain demonstrates the promise that our framework holds for the wide range of multi-agent domains for which AA is critical.

Acknowledgments

This research was supported by DARPA award No. F30602-98-2-0108 (Control of Agent-Based Systems) and managed by ARFL/Rome Research Site.

References

[1] Chalupsky, H., Gil, Y., Knoblock, C. A., Lerman, K., Oh, J., Pynadath, D. V., Russ, T. A., and Tambe, M. Electric elves: Applying agent technology to support human organizations. In *Proc. of the IAAI. Conf.*, 2001.

[2] Collins, J., Bilot, C., Gini, M., and Mobasher, B. Mixed-init. dec.-supp. in agent-based auto. contracting. In *Proc. of the Conf. on Auto. Agents*, 2000.

[3] Dorais, G. A., Bonasso, R. P., Kortenkamp, D., Pell, B., and Schreckenghost, D. Adjustable autonomy for human-centered autonomous systems on mars. In *Proc. of the Intn'l Conf. of the Mars Soc.*, 1998.

[4] Ferguson, G., Allen, J., and Miller, B. TRAINS-95 : Towards a mixed init. plann. asst. In *Proc. of the Conf. on Art. Intell. Plann. Sys.*, pp. 70–77.

[5] Horvitz, E., Jacobs, A., and Hovel, D. Attention-sensitive alerting. In *Proc. of the Conf. on Uncertainty and Art. Intell.*, pp. 305–313, 1999.

[6] Lesser, V., Atighetchi, M., Benyo, B., Horling, B., Raja, A., Vincent, R., Wagner, T., Xuan, P., and Zhang, S. X. A multi-agent system for intelligent environment control. In *Proc. of the Conf. on Auto. Agents*, 1994.

[7] Mitchell, T., Caruana, R., Freitag, D., McDermott, J., and Zabowski, D. Exp. with a learning personal asst. *Comm. of the ACM*, 37(7):81–91, 1994.

[8] Puterman, M. L. *Markov Decision Processes*. John Wiley & Sons, 1994.

[9] Quinlan, J. R. *C4.5: Progs. for Mach. Learn.* Morgan Kaufmann, 1993.

[10] Scerri, P., Pynadath, D. V., and Tambe, M. Adjustable autonomy in real-world multi-agent environments. In *Proc. of the Conf. on Auto. Agents*, 2001.

[11] Tambe, M., Pynadath, D. V., Chauvat, N., Das, A., and Kaminka, G. A. Adaptive agent integration architectures for heterogeneous team members. In *Proc. of the Intn'l Conf. on MultiAgent Sys.*, pp. 301–308, 2000.

[12] Tollmar, K., Sandor, O., and Schōmer, A. Supp. soc. awareness: @Work design & experience. In *Proc. of the ACM Conf. on CSCW*, pp. 298–307, 1996.

Chapter 13

BUILDING EMPIRICALLY PLAUSIBLE MULTI-AGENT SYSTEMS

A Case Study of Innovation Diffusion

Edmund Chattoe
Department of Sociology, University of Oxford

Abstract Multi-Agent Systems (MAS) have great potential for explaining interactions among heterogeneous actors in complex environments: the primary task of social science. I shall argue that one factor hindering realisation of this potential is the neglect of *systematic* data use and appropriate data collection techniques. The discussion will centre on a concrete example: the properties of MAS to model innovation diffusion.

1. Introduction

Social scientists are increasingly recognising the potential of MAS to cast light on the central conceptual problems besetting their disciplines. Taking examples from sociology, MAS is able to contribute to our understanding of emergence [11], relations between micro and macro [4], the evolution of stratification [5] and unintended consequences of social action [9]. However, I shall argue that this potential is largely unrealised for a reason that has been substantially neglected: the relation between data collection and MAS design. I shall begin by discussing the prevailing situation. Then I shall describe a case study: the data requirements for MAS of innovation diffusion. I shall then present several data collection techniques and their appropriate contribution to the proposed MAS. I shall conclude by drawing some more general lessons about the relationship between data collection and MAS design.

2. Who Needs Data?

At the outset, I must make two exceptions to my critique. The first is to acknowledge the widespread instrumental use of MAS. Many computer scientists studying applied problems do not regard data collection about social behaviour as an important part of the design process. Those interested in co-operating robots on a production line assess simulations in instrumental terms. Do they solve the problem in a timely robust manner?

The instrumental approach cannot be criticised provided it only does what it claims to do: solve applied problems. Nonetheless, there is a question about how many meaningful problems are "really" applied in this sense. In practice, many simulations cannot solve a problem "by any means", but have additional constraints placed on them by the fact that the real system interacts with, or includes, humans. In this case, we cannot avoid considering how humans do the task.

Even in social science, some researchers, notably Doran [8] argue that the role of simulation is not to describe the social world but to explore the logic of *theories*, excluding ill-formed possibilities from discussion. For example, we might construct a simulation to compare two theories of social change in industrial societies. Marxists assert that developing industrialism inevitably worsens the conditions of the proletariat, so they are obliged to form a revolutionary movement and overthrow the system. This theory can be compared with a liberal one in which democratic pressure by worker parties obliges the powerful to make concessions.[1] Ignoring the practical difficulty of constructing such a simulation, its purpose in Doran's view is not to describe how industrial societies actually change. Instead, it is to see whether such theories are capable of being formalised into a simulation generating the right outcome: "simulated" revolution or accommodation. This is also instrumental simulation, with the pre-existing specification of the social theory, rather than actual social behaviour, as its "data".

Although such simulations are unassailable on their own terms, their relationship with data also suggests criticisms in a wider context. Firstly, is the rejection of ill-formed theories likely to narrow the field of possibilities very much? Secondly, are existing theories sufficiently well focused and empirically grounded to provide useful "raw material" for this exercise? Should we just throw away all the theories and start again?

The second exception is that many of the most interesting social simulations based on MAS *do* make extensive use of data [1, 16]. Nonetheless, I think it is fair to say that these are "inspired by" data rather than based on it. From my own experience, the way a set of data gets turned into a simulation is something of a "dark art" [5]. Unfortunately, even simulation inspired by data is untypical. In practice, many simulations are based on agents with BDI architectures (for example) not because empirical evidence suggests that people think like this but because the properties of the system are known and the programming is manageable. This approach has unfortunate consequences since the designer has to measure the parameters of the architecture. The BDI architecture might involve decision weights for example and it must be possible to measure these. If, in fact, real agents do not make decisions using a BDI approach, they will have no conception of weights and these will not be measurable or, worse, unstable artefacts of the measuring technique. Until they have been measured, these entities might be described as “theoretical” or “theory constructs”. They form a coherent part of a theory, but do not necessarily have any meaning in the real world.

Thus, despite some limitations and given the state of “normal science” in social simulation, this chapter can be seen as a thought experiment. Could we build MAS genuinely “based on” data? Do such MAS provide better understanding of social systems and, if so, why?

3. The Case Study: Innovation Diffusion

Probably the best way of illustrating these points is to choose a social process that has not yet undergone MAS simulation. Rogers [18] provides an excellent review of the scope and diversity of innovation diffusion research: the study of processes by which practices spread through populations. Despite many excellent qualitative case studies, “normal science” in the field still consists of statistical curve fitting on retrospective aggregate data about the adoption of the innovation.

Now, by contrast, consider innovation diffusion from a MAS perspective. Consider the diffusion of electronic personal organisers (EPO). For each agent, we are interested in all message passing, actions and cognitive processing which bears on EPO purchase and use. These include seeing an EPO in use or using one publicly, hearing or speaking about its attributes (or evaluations of it), thinking privately about its relevance to existing practices (or pros and cons relative to other solutions), having it demonstrated (or demonstrating it). In addition, individuals may discover or recount unsatisfied “needs” which are (currently or subsequently) seen to match EPO attributes, they may actually buy an EPO or seek more information.

A similar approach can be used when more “active” organisational roles are incorporated. Producers modify EPO attributes in the light of market research and technical innovations. Advertisers present them in ways congruent with prevailing beliefs and fears: “inventing” uses, allaying fears and presenting information. Retailers make EPO widely visible, allowing people to try them and ask questions.

This approach differs from the traditional one in two ways. Firstly, it is explicit about relevant social processes. Statistical approaches recognise that the number of new adopters is a function of the number of existing adopters but “smooth over” the relations between different factors influencing adoption. It is true that if all adopters are satisfied, this will lead to further adoptions through demonstrations, transmission of positive evaluations and so on. However, if some are not, then the outcome may be unpredictable, depending on distribution of satisfied and dissatisfied agents in social networks. Secondly, this approach involves almost no theoretical terms in the sense already defined. An ordinary consumer could be asked directly about any of the above behaviours: "Have you ever seen an EPO demonstrated?" We are thus assured of measurability right at the outset.

The mention of social networks shows why questions also need to be presented spatially and temporally. We need to know not just whether the consumer has exchanged messages, but with whom and when. Do consumers first collect information and then make a decision or do these tasks in parallel?

The final (and hardest) set of data to obtain concerns the cognitive changes resulting from various interactions. What effect do conversations, new information, observations and evaluations have? Clearly this data is equally hard to collect in retrospect - when it may not be recalled - or as it happens - when it may not be recorded. Nonetheless, the problem is with elicitation not with the nature of the data itself. There is nothing theoretical about the question "What did you think when you first heard about EPO?"

I hope this discussion shows that MAS are actually very well suited to "data driven" development because they mirror the "agent based" nature of social interaction. Paradoxically, the task of calibrating them is easier when architectures are less dependent on categories originating in theory rather than everyday experience. Nonetheless, a real problem remains. The "data driven" MAS involves data of several different kinds that must be elicited in different ways. Any single data collection technique is liable not only to gather poor data outside its competence but also to skew the choice of architecture by misrepresenting the key features of the social process.

4. Data Collection Techniques

In this section, I shall discuss the appropriate role of a number of data collection techniques for the construction of a "data driven" MAS.

Surveys [7]: For relatively stable factors, surveying the population may be effective in discovering the distribution of values. Historical surveys can also be used for exogenous factors (prices of competing products) or to explore rates of attitude change.

Biographical Interviews [2]: One way of helping with recall is to take advantage of the fact that people are much better at remembering "temporally organised" material. Guiding them through the "history" of their own EPO adoption may be more effective than asking separate survey questions. People may "construct" coherence that was not actually present at the time and there is still a limit to recall. Although interviewees should retain general awareness of the *kinds* of interactions influential in decision (and clear recall of "interesting" interactions), details of number, kind and order of interactions may be lost.

Ethnographic Interviews [12]: Ethnographic techniques were developed for elicitation of world-views: terms and connections between terms constituting a subjective frame of reference. For example, it may not be realistic to assume an objective set of EPO attributes. The term "convenient" can depend on consumer practices in a very complex manner.

Focus Groups [19]: These take advantage of the fact that conversation is a highly effective elicitation technique. In an interview, accurate elicitation of EPO adoption history relies heavily on the perceptiveness of the interviewer. In a group setting, each respondent may help to prompt the others. Relatively "natural" dialogue may also make respondents less self-conscious about the setting.

Diaries [15]: These attempt to solve recall problems by recording relevant data at the time it is generated. Diaries can then form the basis for further data collection, particularly detailed interviews. Long period diaries require highly motivated respondents and appropriate technology to "remind" people to record until they have got into the habit.

Discourse and Conversation Analysis [20, 21]: These are techniques for studying the organisation and content of different kinds of information exchange. They are relevant for such diverse sources as transcripts of focus groups, project development meetings, newsgroup discussions and advertisements.

Protocol Analysis [17]: Protocol analysis attempts to collect data in more naturalistic and open-ended settings. Ranyard and Craig present subjects with "adverts" for instalment credit and ask them to talk about the choice. Subjects can ask for information. The information they ask for and the order of asking illuminate the decision process.

Vignettes [10]: Interviewees are given naturalistic descriptions of social situations to discuss. This allows the exploration of counter-factual conditions: what individuals might do in situations that are not observable. (This is particularly important for new products.) The main problems are that talk and action may not match and that the subject may not have the appropriate experience or imagination to engage with the vignette.

Experiments [14]: In cases where a theory is well defined, one can design experiments that are analogous to the social domain. The common problems with this approach is ecological validity - the more parameters are controlled, the less analogous the experimental setting. As the level of control increases, subjects may get frustrated, flippant and bored.

These descriptions don't provide guidance for practical data collection but that is not the intention. The purpose of this discussion is threefold. Firstly, to show that data collection methods *are* diverse: something often obscured by methodological preconceptions about "appropriate" techniques. Secondly, to suggest that different techniques are appropriate to different aspects of a "data driven" MAS. Few aspects of the simulation discussed above are self-evidently ruled out from data collection. Thirdly, to suggest that prevailing data poor MAS may have more to do with excessive theory than with any intrinsic problems in the data required.

There are two objections to these claims. Firstly, all these data collection methods have weaknesses. However, this does not give us grounds for disregarding them: the weakness of inappropriately collected data (or no data at all) is clearly greater. It will be necessary to triangulate different techniques, particularly for aspects of the MAS which sensitivity analysis shows are crucial to aggregate outcomes. The second "difficulty" is the scale of work and expertise involved in building "data driven" MAS. Even for a simple social process, expertise may be required in several data collection techniques. However, this difficulty is intrinsic to the subject matter. Data poor MAS may choose to ignore it but they do not resolve it.

5. Conclusions

I have attempted to show two things. Firstly, MAS can be used to model social processes in a way that avoids theoretical categories Secondly, different kinds of data for MAS *can* be provided by appropriate techniques. In the conclusion, I discuss four general implications of giving data collection "centre stage" in MAS design.

Dynamic Processes: MAS draws attention to the widespread neglect of process in social science.[2] Collection of aggregate time series data does little to *explain* social change even when statistical regularities can be established. However, attempts to base genuinely dynamic models (such as MAS) on data face a fundamental problem. There is no good time to ask about a dynamic process. Retrospective data suffers from problems with recall and rationalisation. Prospective data suffers because subjects cannot envisage outcomes clearly and because they cannot assess the impact of knowledge they haven't yet acquired. If questions are asked at more than one point, there are also problems of integration. Is the later report more accurate because the subject knows more or less accurate because of rationalisation? Nonetheless, this problem is again intrinsic to the subject matter and ignoring it will not make it go away. Triangulation of methods may address the worst effects of this problem but it needs to be given due respect.

Progressive Knowledge: Because a single research project cannot collect all the data needed for even a simple "data driven" MAS, progressive production and effective organisation of knowledge will become a priority. However, this seldom occurs in social science (Davis 1994). Instead data are collected with particular theory constructs in mind, rendering them unsuitable for reuse. To take an example, what is the role of “conversation” in social networks? Simulation usually represents information transmission through networks as broadcasting of particulate information. In practice, little information transmission is unilateral or particulate. What impact does the fact that people converse have on their mental states? We know about the content of debates (discourse analysis) and the dynamics of attitudes (social psychology) but almost nothing about the interaction between the two.

Data Collection as a Design Principle: Proliferation of MAS architectures suggests that we need to reduce the search space for social simulation. In applied problems, this is done by pragmatic considerations: cost, speed and “elegance”. For descriptive simulations, the ability to collect data may serve a corresponding role. It is always worth asking why MAS need unobtainable data. The reasons may be pragmatic but if they are not, perhaps the architecture should be made less dependent on theoretical constructs so it can use data already collected for another purpose.

Constructive Ignorance: The non-theoretical approach also suggests important research questions obscured by debates over theoretical constructs. For example, do people transmit evaluations of things they don’t care about? What is the impact of genuine dialogue on information transmission? When does physical distance make a difference to social network structure? Answers to these questions would be useful not just for innovation diffusion but in debates about socialisation, group formation and stratification. Formulating questions in relatively non-theoretical terms also helps us to see what data collection techniques might be appropriate. Recognising our ignorance (rather than obscuring it in abstract debates about theory constructs) also helps promote a healthy humility!

In conclusion, focusing MAS design on data collection may not resolve the difficulties of understanding complex systems, but it definitely provides a novel perspective for their examination.

Notes

[1] This example illustrates the meaning of “theory” in social science. A theory is a set of observed regularities (revolutions) explained by postulated social processes (exploitation of the proletariat, formation of worker groups, recognition that revolution is necessary).

[2] The problem has recently been recognised (Hedström and Swedburg 1998) but the role of simulation in solving it is still regarded with scepticism by the majority of social scientists.

References

[1] Bousquet, F. et al. Simulating Fishermen’s Society, In: Gilbert, N. and Doran, J. E. (Eds.) *Simulating Societies* London: UCL Press, 1994.

[2] Chamberlayne, P., Bornat, J. and Wengraf, T. (eds.) *The Turn to Biographical Methods in the Social Sciences: Comparative Issues and Examples* London: Routledge, 2000.

[3] Chattoe, E. Why Is Building Multi-Agent Models of Social Systems So Difficult? A Case Study of Innovation Diffusion, XXIV International Conference of Agricultural Economists IAAE, Mini-Symposium on Integrating Approaches for Natural Resource Management and Policy Analysis, Berlin, 13–19 August, 2000.

[4] Chattoe, E. and Heath, A. A New Approach to Social Mobility Models: Simulation as "Reverse Engineering" Presented at the BSA Conference, Manchester Metropolitan University, 9-12 April, 2001.

[5] Chattoe, E. and Gilbert, N. A Simulation of Adaptation Mechanisms in Budgetary Decision Making, in Conte, R. et al. (Eds.) *Simulating Social Phenomena* Berlin: Springer-Verlag, 1997.

[6] Davis, J. A. What's Wrong with Sociology? *Sociological Forum*, 9:179-197, 1994.

[7] De Vaus, D. A. *Surveys in Social Research*, 3rd ed. London: UCL Press, 1991.

[8] Doran J. E. From Computer Simulation to Artificial Societies, *Transactions of the Society for Computer Simulation*, 14:69-77, 1997.

[9] Doran, J. E. Simulating Collective Misbelief *Journal of Artificial Societies and Social Simulation*, 1(1), <http://www.soc.surrey.ac.uk/JASSS/1/1/3.html>, 1998.

[10] Finch, J. The Vignette Technique in Survey Research, *Sociology*, 21:105-114, 1987.

[11] Gilbert, N. Emergence in Social Simulation, In Gilbert, N. and Conte, R. (eds.) *Artificial Societies*. London: UCL Press, 1995.

[12] Gladwin, C. H. *Ethnographic Decision Tree Modelling*, Sage University Paper Series on Qualitative Research Methods Vol. 19 London: Sage Publications, 1989.

[13] Hedström, P. and Swedberg, R. *Social Mechanisms: An Analytical Approach to Social Theory* Cambridge: CUP, 1998.

[14] Hey, J. D. *Experiments in Economics* Oxford: Basil Blackwell, 1991.

[15] Kirchler, E. Studying Economic Decisions Within Private Households: A Critical Review and Design for a "Couple Experiences Diary", *Journal of Economic Psychology*, 16:393–419, 1995.

[16] Moss, S. Critical Incident Management: An Empirically Derived Computational Model, *Journal of Artificial Societies and Social Simulation*, 1(4), <http://www.soc.surrey.ac.uk/JASSS/1/4/1.html>, 1998.

[17] Ranyard, R. and Craig, G. Evaluating and Budgeting with Instalment Credit: An Interview Study, *Journal of Economic Psychology*, 16:449–467, 1995.

[18] Rogers, E. M. *Diffusion of Innovations*, 4th ed. New York: The Free Press, 1995.

[19] Wilkinson, S. Focus Group Methodology: A Review, *International Journal of Social Research Methodology*, 1:181-203, 1998.

[20] Wood, L. A. and Kroger, R. O. *Doing Discourse Analysis: Methods for Studying Action in Talk and Text* London: Sage Publications, 2000.

[21] Wooffitt, R. and Hutchby, I. *Conversation Analysis* Cambridge: Polity Press, 1998.

Chapter 14

ROBOTIC PLAYMATES

Analysing Interactive Competencies of Children with Autism Playing with a Mobile Robot

Kerstin Dautenhahn[1], Iain Werry[2], John Rae[3], Paul Dickerson[3],
Penny Stribling[3], and Bernard Ogden[1]
[1]*University of Hertfordshire,* [2]*University of Reading,* [3]*University of Surrey Roehampton*

Abstract This chapter discusses two analysis techniques that are being used in order to study how children with autism interact with an autonomous, mobile and 'social' robot in a social setting that also involves adults. A quantitative technique based on micro-behaviours is outlined. The second technique, Conversation Analysis, provides a qualitative and more detailed investigation of the sequential order, local context and social situatedness of interaction and communication competencies of children with autism. Preliminary results indicate the facilitating role of the robot and its potential to be used in autism therapy.

1. The Aurora Project

Computers, virtual environments and robots (e.g. [15], [9]) are increasingly used as interactive learning environments in autism therapy[1]. Since 1998 the Aurora project has studied the development of a mobile, autonomous and 'social robot' as a therapeutic tool for children with autism, see e.g. [1] for more background information. Here, the context in which robot-human interactions occur is deliberately playful and 'social' (involving adults). In a series of trials with 8-12 year-old autistic children we established that generally children with autism enjoy interacting with the robotic toy, and show more engaging behaviour when playing with the robot as opposed to a non-interactive toy [16], [17]. Also, the role of the robot as a social mediator was investigated in trials with pairs of autistic children. Results showed a spectrum of social and non-social play and communication that occurred in robot-child and child-

child interactions [18]. Overall, results so far seem to indicate that a) the robot can serve as an interesting and responsive interaction partner (which might be used in teaching social interaction skills), and b) that the robot can potentially serve as a social facilitator and a device that can be used to assess the communication and social interaction competencies of children with autism. In order to investigate robot-human interactions systematically, in the Aurora project two analysis techniques have been developed and tested.

2. Analysis of Interactions

2.1 Methodological Issues

Trials are conducted at a room at Radlett Lodge School - the boarding school that the children participating in the trial attend. This has many advantages such as familiar surroundings for the children and the availability of teachers who know the children well. The fact that the children do not need to travel and that the trials inflict a minimum amount of disruption to lessons also helps the children to adapt to the change in schedule.

The room used is approximately two meters by three meters, and is set aside for us and so does not contain extra features or excess furniture. The robotic platform used in this research is a Labo-1 robot. The robot is 30cm wide by 40cm long and weighs 6.5kg. It is equipped with eight infrared sensors (four at the front, two at the rear and one at either side), as well as a heat sensor on a swivel mount at the front of the robot. Using its sensors, the robot is able to avoid obstacles and follow a heat source such as a child. Additionally, a speech synthesiser unit can produce short spoken phrases using a neutral intonation. The robot is heavy enough to be difficult for the children to pick up and is robust enough to survive an average trial, including being pushed around. The programming of the robot allows it to perform basic actions, such as avoiding obstacles, following children and producing speech. The robot will try to approach the child, respond vocally to his presence, and avoid other obstacles - as well as not coming into actual contact with the child. All trials are videotaped. In the following, the quantitative approach described in section 2.2 analyses robot-human interactions in comparative trials. Section 2.3 introduces a qualitative approach that is applied to analyse the interactions of one child with the robot and adults present during the trials.

2.2 A Quantitative Approach

The trials involve the child showing a wide variety of actions and responses to situations. Unexpected actions are usually positive results and free expression and full-body movements are encouraged. In order to examine the interactions and evaluate the robot's interactive skills we developed a quantitative method of analysing robot-human interactions, based on a method used previously to analyse child-adult interactions[2].

This section describes the analysis of robot-human interactions in a comparative study where seven children interact separately with the mobile robot and a non-interactive toy[3]. Trials are conducted in three sections. The first section involves the child interacting with a toy truck, approximately the same size as the robotic platform. The second section consists of both the toy truck and the robotic platform present simultaneously whereby the robot is switched off. The third section involves the robot without the toy truck, see figure 14.1. In half the trials the order of the first and last section is reversed. This structure allows us to compare interactions with the robot with those of a solely passive object. Timing of the sections vary, typically the first and third section are four minutes while the second section is two minutes, depending on the enjoyment of the child.

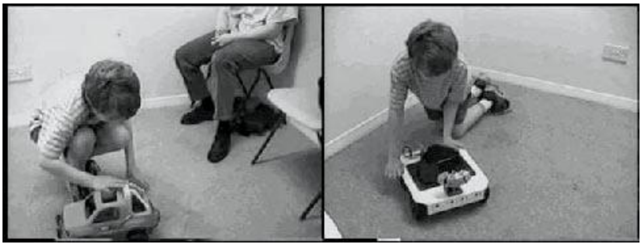


Figure 14.1. Ivan playing with the toy truck (left) and the robot (right). All names of children used in this chapter are pseudonyms.

The trial video is segmented into one-second intervals, and each second is analysed for the presence of various behaviours and actions by the child (after [14], with criteria altered for our particular application). Trials are analysed using a set of fourteen criteria, which are broken into two general categories. The first category consists of the criteria eye gaze, eye contact, operate, handling, touch, approach, move away and attention. This category depends on a focus of the action or behaviour and this focus further categorises the analysis of the behaviour. The second category consists of the criteria vocalisation, speech, verbal stereotype, repetition and blank. The focus of these actions are recorded where possible.

The histogram in figure 14.2 shows a sample of the results of trials using this analysis method, focused on the criterium *eye gaze*. As can be seen, the values for gaze are considerably higher when focused on the robot than the toy truck for three of the seven children shown (Ivan, Oscar, Peter). Adam looked at the

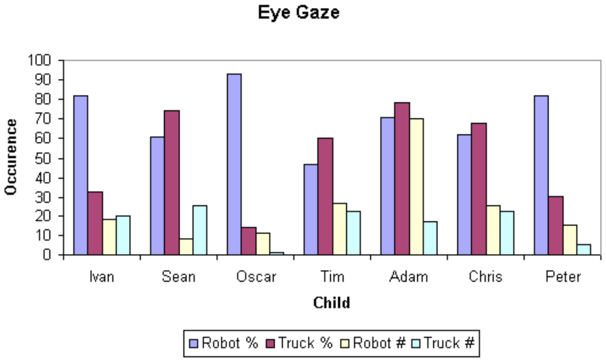


Figure 14.2. Eye gaze behaviours of seven children who interacted with the interactive robot and a passive toy truck in a comparative study. Shown is the percentage of time during which the behaviour occurred in the particular time interval analysed (%), as well as the number of times the behaviour was observed (#). Note, that the length of the trial sections can vary.

robot very frequently but briefly. Chris, Sean and Tim direct slightly more eye gaze behaviour towards the toy truck. The quantitative results nicely point out individual differences in how the children interact with the robot, data that will help us in future developments. Future evaluations with the full list of criteria discussed above will allow us to characterise the interactions and individual differences in more detail.

2.3 A Qualitative Approach

This section considers the organisation of interaction in the social setting that involves the child, the robot and adults who are present. The following analysis draws on the methods and findings of Conversation Analysis (CA) an approach developed by Harvey Sacks and colleagues (e.g. [13]) to provide a systematic analysis of everyday and institutional talk-in-interaction. Briefly, CA analyses the fine details of naturalistic talk-in-interaction in order to identify the practices and mechanisms through which sequential organisation, social design and turn management are accomplished. For overviews and transcription conventions see [5], [11]. This requires an inductive analysis that reaches beyond the scope of quantitative measures of simple event frequency. A basic principle of CA is that turns at talk are *"context-shaped and context-renewing"* ([4], p. 242). This has a number of ramifications, one of which is that the action performed by an utterance can depend on not just what verbal or other elements it consists of, but also its *sequential location*. Consider for example how a greeting term such as "hello" is unlikely to be heard as "doing a greeting" unless it occurs in a specific location, namely in certain opening turns in an interaction ([12], vol. 2, p.36, p.188).

It is the capacity to address the *organisation of embodied action*, which makes CA particularly relevant for examining robot-child interactions. In addition to examining vocal resources for interaction, CA has also been applied to body movement (in a somewhat different way to the pioneering work of Kendon, [8]), e.g. [3]). It has also been applied to interactions with, or involving, non-human artifacts (such as computers [2]). We aim to provide a brief illustration of the relevance of CA to examining both the interactional competencies of children with autism and their interactions with the robot by sketching some details from a preliminary analysis of an eight minute session involving one boy, Chris (C), the robot (R) and a researcher (E).

Whilst pragmatic communicative competence is not traditionally attributed to people with autism (indeed the iconic image of the Autist is that of being isolated and self-absorbed) attention to the autistic child's activities in their interactional context can reveal communicative competence which might otherwise be missed. It can be established that when the context is considered, many of Chris's actions (vocal and non-vocal) can be seen to be responsive to things that the robot does. For example at one point Chris emits a surprised exclamation "oooh!". Extract 1 in figure 14.3 shows that this is evidently responsive to a sudden approach from the robot.

```
Extract 1 [T&R Chris 17:4-5]
Note: double parentheses show comments or descriptions.
1  R: ****** ((motor noise as R rapidly approaches C))
2->C: oooh!

Extract 2 [T&R Chris 4:20-5:9]
Note: dashes in parentheses show silences, each dash is 0.1sec.
C moves and is tracked by R, R lines up with C then:
1  R: ther:e you are. ther:e you are.((R approaches C))
2->   (-----) ((R goes past C))
3->C: not very good at ^steering it[s:el:f
4  E:                              [(hah)(huh)

Extract 3 [T&R Chris 13:6-12]
1  E: yeh that tells you what the programme is
2  C: (ys) an it <an it tells you what moo::d it's in
3     (---------+-----)  ((C moves to the right of R))
                |
             ((R moves to the point just vacated by C))
4  R: do do
5     (---------+) ((R rotates to the right to face C))
6  C: an it's in a (fo'owing) moo::d
7  E: (hah)
8  C: it's in a following moo::d
9  E: yes it is ((smiley voice))
```

Figure 14.3. Extracts of transcriptions.

This attention to sequential organisation can provide a refreshing perspective on some of the 'communication deficits' often thought characteristic of autism. For example, ‘Echolalia’ [7], (which can be immediate or delayed) is typically conceptualised as talk which precisely reproduces, or echoes, previously overheard talk constituting an inappropriate utterance in the assumed communicative context. Likewise ‘Perservation’ or inappropriate topic maintenance is also understood as a symptom of autism. Despite more recent developments that have considered echolalia’s capacity to achieve communicative goals [10] and have raised the potential relevance of conversation analysis in exploring this issue [19] the majority of autism researchers treat the echolalic or perservative talk of children with autism as symptomatic of underlying pathology.

In our data Chris makes ten similar statements about the robot’s poor steering ability such as “not very good at ^steering its:el:f ”. In a content analysis even a quite specific category ‘child comments on poor steering ability’ would pull these ten utterances into a single category leaving us likely to conclude that Chris’s contribution is ‘perseverative’ or alternatively ‘delayed-echolalic’. However a CA perspective provides a more finely honed approach allowing us to pay attention to the distinct form of each utterance, its specific embedding in the interactional sequence and concurrent synchronous movement and gesture. For example extract 2 in figure 14.3 shows how one of Chris’s “not very good at ^steering it[s:el:f” statements (line 3) is clearly responsive to the robot approaching, but going past him (line 2).

Chris also makes seven, apparently repetitious, statements about the robot being in a certain “mood” in the course of a 27 second interval. Three of these are shown in Extract 3 in figure 14.3 (in lines 2, 6 and 8). Chris’s utterance in line 2 follows a number of attempts by him to establish that an LCD panel on the back of robot (the “it” in line 2) tells one about the “mood” of the robot (an issue for the participants here apparently being the appropriateness of the term “mood”, as opposed to “programme”). By moving himself (in line 3) and characterising the robot’s tracking movements (from lines 3 - 5) as evidence for the robot being in a “following mood” (line 6) Chris is able to use the robot’s tracking movements as a kind of practical demonstration of what he means when he refers to “mood”. In this way, rather than being an instance of ‘inappropriate’ repetition, the comment about mood (line 6) firstly involves a change from talking about the LCD panel to making a relevant observation about the robot’s immediate behaviour, secondly it apparently addresses an interactionally relevant issue about the meaning of word “mood”. Incidentally, it can be noted that the repetition of line 6 which occurs in line 8 also has good interactional reasons. Line 6 elicits a kind of muted laugh from E – a response that does not demonstrably display E’s understanding of C’s prior utterance. C therefore undertakes self-repair in line 8, repeating his characterisation, and this time securing a fuller response from E “yes it is” (in line 9).

By moving away from studying vocal behaviour in isolation to focusing on embodied action in its sequential environments, CA can show how a person with autism engages in social action and orients to others through both verbal and non-verbal resources. Here, using naturalistic data involving activities generated and supported by a mobile robot we can demonstrate how talk which might be classified as perservation or echolalia by a content analytic approach is in fact a pragmatically skilled, socially-oriented activity. The practical benefit of orientation to interactive context lies in developing our understanding of the exact processes involved in interactions that include people with autism, thereby helping service providers to identify the precise site of communicative breakdowns in order to support focused intervention.

3. Conclusion

This chapter discussed two techniques for analysing interaction and communication of children with autism in trials involving a social robot, work emerging from the Aurora project. Ultimately, different quantitative and qualitative analysis techniques are necessary to fully assess and appreciate the communication and interaction competencies of children with autism. Results will provide us with valuable guidelines for the systematic development of the design of the robot, its behaviour and interaction skills, and the design of the trial sessions.

Acknowledgments

The AURORA project is supported by an EPSRC grant (GR/M62648), Applied AI Systems Inc. and the teaching staff at Radlett Lodge School.

Notes

1. The autistic disorder is defined by specific diagnostic criteria, specified in DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Association, 1994). Individuals with autism show a broad spectrum of difficulties and abilities, and vary enormously in their levels of overall intellectual functioning [6]. However, all individuals diagnosed with autism will show impairments in communication and social interaction skills.

2. The analysis of the videotapes focuses on the child. However, since we are trying to promote social interaction and communication, the presence of other people is not ignored, rather examined from the perspective of the child.

3. Previous results with four children were published in [16], [17].

References

[1] Kerstin Dautenhahn and Iain Werry. Issues of robot-human interaction dynamics in the rehabilitation of children with autism. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. W. Wilson, editors, *Proc. From animals to animats 6, The Sixth International Conference on the Simulation of Adaptive Behavior (SAB2000)*, pages 519–528, 2000.

[2] D. Frohlich, P. Drew, and A. Monk. Management of repair in human-computer interaction. *Human-Computer Interaction*, 9:385–425, 1994.

[3] C. Goodwin. *Conversational organization: interaction between speakers and hearers*. Academic Press, New York, 1981.

[4] J. C. Heritage. *Garfinkel and ethnomethodology*. Polity Press, Cambridge, 1984.

[5] I. Hutchby and R. Wooffitt. *Conversation Analysis: principles, practices and applications*. Polity Press, Cambridge, 1998.

[6] Rita Jordan. *Autistic Spectrum Disorders – An Introductory Handbook for Practitioners*. David Fulton Publishers, London, 1999.

[7] L. Kanner. Irrelevant metaphorical language in early infantile autism. *American Journal of Psychiatry*, 103:242–246, 1946.

[8] A. Kendon. *Conducting Interaction: Patterns of Behaviour in Focused Encounters*. Cambridge University Press, 1990.

[9] F. Michaud, A. Clavet, G. Lachiver, and M. Lucas. Designing toy robots to help autistic children - an open design project for electrical and computer engineering education. *Proc. American Society for Engineering Education*, 2000.

[10] B. M. Prizant and P. J. Rydell. Analysis of functions of delayed echolalia in autistic children. *Journal of Speech and Hearing Research*, 27:183–192, 1984.

[11] G. Psathas. *Conversation analysis: the study of talk-in-interaction*. Sage, London, 1992.

[12] H. Sacks. *Lectures on conversation*. Blackwell, Oxford, UK, 1992.

[13] H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organisation of turn-taking for conversation. *Language*, 50:696–735, 1974.

[14] C. Tardiff, M.-H. Plumet, J. Beaudichon, D. Waller, M. Bouvard, and M. Leboyer. Micro-analysis of social interactions between autistic children and normal adults in semi-structured play situations. *International Journal of Behavioural Development*, 18(4):727–747, 1995.

[15] S. Weir and R. Emanuel. Using LOGO to catalyse communication in an autistic child. Technical report, DAI Research Report No. 15, University of Edinburgh, 1976.

[16] Iain Werry, Kerstin Dautenhahn, and William Harwin. Evaluating the response of children with autism to a robot. Proc. RESNA 2001, Rehabilitation Engineering and Assistive Technology Society of North America, 22-26 June 2001, Reno, Nevada, USA, 2001.

[17] Iain Werry, Kerstin Dautenhahn, and William Harwin. Investigating a robot as a therapy partner for children with autism. Proc. AAATE 2001, 6th European Conference for the Advancement of Assistive Technology (AAATE 2001), 3-6 September 2001 in Ljubljana / Slovenia., 2001.

[18] Iain Werry, Kerstin Dautenhahn, Bernard Ogden, and William Harwin. Can social interaction skills be taught by a social agent? the role of a robotic mediator in autism therapy. In M. Beynon, C. L. Nehaniv, and K. Dautenhahn, editors, *Proc. CT2001, The Fourth International Conference on Cognitive Technology: Instruments of Mind, LNAI 2117*, pages 57–74, Berlin Heidelberg, 2001. Springer-Verlag.

[19] A. J. Wootton. An investigation of delayed echoing in a child with autism. *First Language*, 19:359–381, 2000.

Chapter 15

MOBILE ROBOTIC TOYS AND AUTISM

Observations of Interaction

François Michaud and Catherine Théberge-Turmel
Université de Sherbrooke

Abstract To help children with autism develop social skills, we are investigating the use of mobile robotic toys that can move autonomously in the environment and interact in various manners (vocal messages, music, visual cues, movement, etc.), in a more predictable and less intimidating way. These interactions are designed to build up their self-esteem by reinforcing what they do well. We report tests done with autistic children using different robots, each robot having particular characteristics that allow to create interesting interactions with each child.

1. Introduction

Autism is characterized by abnormalities in the development of social relationships and communication skills, as well as the presence of marked obsessive and repetitive behavior. Despite several decades of research, relatively little is understood about the causes of autism and there is currently no cure for the condition. However education, care and therapeutic approaches can help people with autism maximize their potential, even though impairments in social and communication skills may persist throughout life.

As engineers, we got interested in the idea of designing mobile robotic toys to help children with autism learn to develop appropriate social skills. For an autistic child, a robot may be less intimidating and more predictable than a human. A robot can follow a deterministic play routine and also adapt over time and change the ways it responds to the world, generating more sophisticated interactions and unpredictable situations that can help capture and retain the child's interest. Robotic toys also have the advantage that they can be programmed to respond differently to situations and events over time. This flexibility allows robotic toys to evolve from simple machines to systems that demonstrate more complex behavior patterns.

The general goal is to create learning situations that stimulate children, get them to socialize and integrate them in a group. People with autism are aware that they have difficulties making sense of the outside world. To help them move from predictable, solitary and repetitive situations where they feel safe to socially interact with the world, the first objective of our robotic toys is to build up their self-esteem by reinforcing what they do good. The idea is to ask the child to do something, and to reward the child if the request is successfully satisfied. To make this work, the activities and the rewards must be something that interests the child, and one of the challenges is to get the attention of the child and get them interested in interacting. Another advantage of robotic toys is that they can have special devices that are particular interesting to these children, trying to find incentives to make them open up to their surroundings. Since each child is a distinct individual with preferences and capabilities, we are not seeking to design one complete robotic toy that would work with all autistic children. We want to observe the possible factors that might influence the child's interests in interacting with a robotic toy, like shape, colors, sounds, music, voice, movements, dancing, trajectory, special devices, etc. To do so, different mobile robots have been used in tests ranging from single sessions of a couple of minutes to consecutive use over a five week period, with autistic children or young adults of 7 to 20 years old. This way, our long term goal is to design robotic toys that can take into account the interests, strengths and weaknesses of each child, generate various levels of predictability, and create a more tailored approach for personalized treatment.

2. Mobile Robotic Toys with Autistic Children

Two types of tests have been conducted with autistic children: short sessions at the École du Touret, and using one robot over a five week period with groups of children and young adults at the S.P.E.C. Tintamarre Summer camp.

2.1 Short Sessions

These sessions were held in two rooms: one regular classroom and a 20'x20' room without tables and chairs. Children were allowed to interact freely with the robots. At all time at least one educator was there to introduce the robot to children, or to intervene in case of trouble. Even though these children were not capable of fluent speech, some were able to understand the short messages generated by the robots. Each session lasted around one hour and a half, allowing eight to ten children to play with the robots. No special attention was put on trial length for each child, since our goal was to let all the children of the class play with the robots in the allocated time slot.

As expected, each child had his or her own ways of interacting with the robots. Some remained seated on the floor, looking at the robot and touching it when it came close to them (if the robot moved to a certain distance, some children just stopped looking at the robot). Others moved around, approaching and touching the robots and sometimes showing signs of excitation. It is very hard to generalize the results of these tests since each child is so different. In addition, the mood of some of the children that participated to all of these sessions was not always the same. But one thing that we can say is that the robots surely caught the attention of the children, making them smile, laugh or react vocally. In general, we did not observe particular attention to the front of the robots (e.g., trying to make eye contact), mostly because most of them have devices all around them. To give a more precise evaluation of our tests, we present observations made with some of the robots used in these trials:

Jumbo. This elephant has a moving head and trunk, one pyroelectric sensor and an infrared range sensor. Jumbo is programmed to move toward the child and to stop at a distance of 20 cm. Once close to the child, Jumbo asks the child to touch one of the three buttons associated with pictograms located on its back. LEDs are used at first to help the child locate the right pictogram, but eventually the LEDs are not used. If the child is successful, Jumbo raises its trunk and plays some music (Baby's Elephant Walk or Asterix the Gaulish). If the child is not responding, the robot asks to play and can try to reposition itself in front of the child. Pictograms on the robot can be easily replaced. This robot revealed to be very robust, even though its pyroelectric lenses got damaged too. One child liked to push the robot around when it was not moving, as shown in Figure 15.1, or to make the robot stay close to her if it was moving away. The pictogram game was also very nice, but children were pressing on the pictograms instead of on the buttons. The music played and movements of the trunk were also very appreciated by the children.

Roball. Roball [3] is a spherical robot capable of navigating in all kind of environments without getting stuck somewhere or falling on the side. Interactions can be done using vocal messages and movement patterns like spinning, shaking or pushing. The majority of children were trying to catch Roball, to grab it or to touch the robot. Some even made it spin (but not always when requested by Roball though). One boy, who did not interact much with almost all of the other robots presented, went by himself in order to play with Roball. One of the games he played was to make the robot roll on the floor between his arms, as shown in Figure 15.2, and eventually let it go forward by itself.

C-Pac. C-Pac is a very robust robot that has removable arms and tail. These removable parts use connectors that have different geometrical shape (star, triangle, hexagon). When successfully assembled, the robot thanks the child and rotates by itself. The robot also asks the child to make it dance by pressing its head. The head then becomes illuminated, and music (La Bamba) is played as the robot dances, and this was very much appreciated by children. C-Pac also has a moving mouth, eyes made of LEDs, an infrared range sensor and pyroelectric sensors to stay close to the child. Children learned rapidly how to play with this robot, even understanding by themselves how to assemble the robot, as shown in Figure 15.3. The removable parts became toys on their own. Children were also surprised when they grabbed the robot by its arms or tail, expecting to grab the robot but instead removing the part from the robot. Note however that the pyroelectric lenses got damaged by the children, and one even took off the plastic cup covering one eye of the robot and tried to ate it.

Bobus. Extremely robust, this robot can detect the presence of a child using pyroelectric sensors. It then slowly moves closer to the child, and when close enough it does simple movements and plays music. Simple requests (like touching) are made to the child and if the child responds at the appropriate time, light effects are generated using the LEDs all around the 'neck' of the robot, and the small ventilator on its head is activated. Very robust, this robot is the only one with pyroelectric senses that did not get damaged. Two little girls really liked the robot, enjoying the light effects, the moving head with the ventilator, and the different textures. Figure 15.4 illustrates one of these girls showing signs of excitation when playing with Bobus. At one point, one girl lifted the robot and was making it roll on its side on top of her legs. She then put the robot on the floor and was making it roll on its side using her legs again, but by lying on top of the robot.

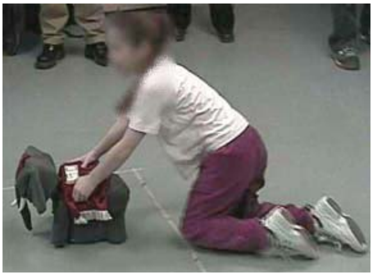


Figure 15.1. Pushing Jumbo around the play area.

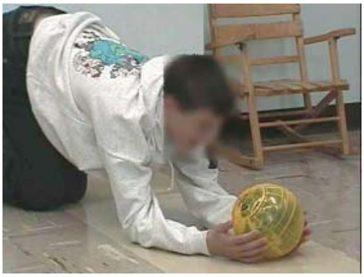


Figure 15.2. Rolling game with Roball.

One very interesting observation was made with a 10 years old girl. When she enters the recreation room, she starts right away to follow the walls, and

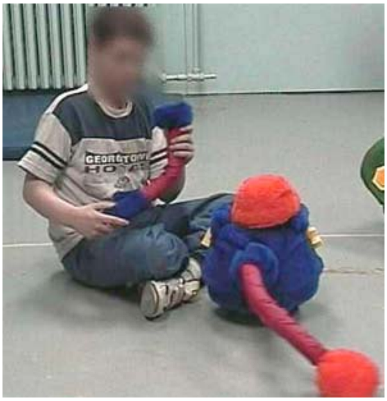


Figure 15.3. Assembling the arms and tail of C-Pac.

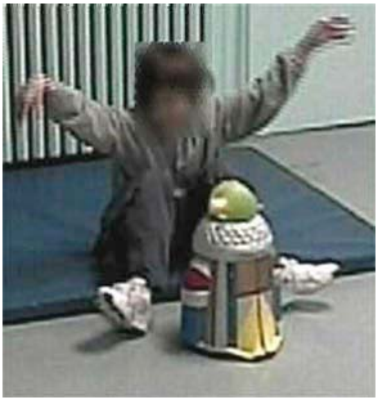


Figure 15.4. Girl showing signs of interest toward Bobus.

she can do this over and over again, continuously. At first, a robot was placed near a wall, not moving. The little girl started to follow the walls of the room, and interacted with the robot for short amounts of time, at the request of the educator as she went by the robot. Eventually, the robot moved away from the walls and she slowly started to stop, first at one particular corner of the room, and then at a second place, to look at the robot moving around. At one point when the robot got to a corner of the room, she changed path and went out of her way to take the robot by its tail and to drag it back to the center of the room where she believed the robot should be. She even smiled and made eye contact with some of us, something that she did not do with strangers. This showed clear indications that having the robot moved in the environment helped her gradually open up to her surroundings.

2.2 Trials at S.P.E.C. Tintamarre Summer Camp

In these trials, Jumbo was used one day a week over a period of five weeks, for 30 to 40 minutes in four different groups. Children and young adults were grouped according to the severity of their conditions, their autonomy and their age. Four to ten people were present in each group, along with two or three educators, and each group had its own room. Children were placed in a circle, sitting down on the floor or on small cubes depending on their physical capabilities. The robot always remained on the floor, and each child played in turns with the pictograms. Once a turn was completed, a new set of pictograms was used.

With the groups that did not have physical disabilities, children manifested their interests as soon as Jumbo entered the room, either by looking at the robot or by going to touch it, to push it, to grab the trunk or by pressing on the pictograms. The music and the dance were very much appreciated by the children. The amount of interactions varied greatly from one child to another. Some remained seated on the floor and played when the robot was close to them. Others either cleared the way in front of the robot, or moved away from its path when it was coming in their direction. The amount of time they remained concentrated on the robot was longer than for the other activities they did as a group. One little girl who did not like animals, had no trouble petting Jumbo. She was also playing in place of others when they took too much time responding to a request or did mistakes. One boy did the same thing (even by going through the circle), and he was very expressive (by lifting his arms in the air) when he succeeded with the pictograms.

To the group of teenagers, Jumbo is real. They talked to the robot, reacted when it was not behaving correctly or when it was not moving toward them. Some educators were also playing along because they were talking to Jumbo as if it was a real animal, by calling its name, asking it to come closer. When Jumbo did not respond correctly and was moving away, educators would say something like "Jumbo! You should clean your ears!" or "Jumbo has big ears but cannot hear a thing!". One boy showed real progress in his participation, his motivation and his interactions because of the robot. His first reaction was to observe the robot from a distance, but he rapidly started to participate. His interest toward the robot was greater than the other kids. He remembered the pictograms and the interactions they had with the robot from one week to another. He also understood how to change the pictograms and asked frequently the educators to let him do it. Another boy also liked to take Jumbo in his arms, like an animal. He showed improvements in shape and color recognition.

3. Discussion

Our tests revealed that autistic children are interested by the movements made by the robots, and enjoy interacting with these devices. Note that it should never be expected that a child will play as intended with the robot. This is part of the game and must be taken into consideration during the design stage of these robots. In that regard, robustness of the robots is surely of great importance, as some of the more fragile designs got damaged, but mostly by the same child. Having removable parts is good as long as they are big enough: all small components or material that can be easily removed should be avoided. Having the robots behave in particular ways (like dancing, playing music, etc.) when the child responds correctly to requests made by the robot becomes a powerful incentive for the child to continue playing with the robots. The idea is to create rewarding games that can be easily understood (because of its simplicity or because it exploit skills developed in other activities like the use of pictograms or geometrical shapes) by the child.

In future tests and with the help of educators, we want to establish a more detailed evaluation process in order to assess the impact of the mobile robotic toys on the development of the child. We also want to improve the robot designs and to have more robots that can be lent to schools over longer periods of time. The robots should integrate different levels of interaction with the child, starting with very simple behaviors to more sophisticated interplay situations. Catching and keeping their attention are important if we want the children to learn, and the observations made with the robots described in the previous section can be beneficial. The idea is not as much as using the robot to make children learn to recognize for instance pictograms (they learn to do this in other activities), but to make them develop social skills like concentration, sharing, turn passing, adaptation to changes, etc. Finding the appropriate reward that would make the child want to respond to the robot's request is very important. Predictability in the robot's behavior is beneficial to help them understand what is going on and how to receive rewards. Also, since the robot is a device that is programmed, the robot's behavior can evolve over time, changing the reinforcing loop over time, to make them learn to deal with more sensory inputs and unpredictability. Finally, to adapt mobile robot toys to each child, reconfigurable robots, using different hardware and software components, might be one solution to explore.

Using interactive robotic toys is surely an interesting idea that has the potential of providing an additional intervention method to the rehabilitation process of autistic children. We are not alone working on this aspect. The AURORA project (AUtonomous RObotic platform as a Remedial tool for children with Autism) [2, 1, 5] is one of such initiatives addressed in the previous chapter.

We are very much encouraged by the observations made, and we will continue to design new mobile robots [4] and to do tests with autistic children. The basic challenge is to design a robot that can catch their attention and help them develop their social skills by building up their self-esteem. At this point, we still need to work on simple ways of interacting with the child, to help them understand how the robot works and exploit the knowledge and skills they acquire in other pedagogical activities. Our hope is that mobile robotic toys can become efficient therapeutic tools that will help children with autism develop early on the necessary social skills they need to compensate for and cope with their disability.

Acknowledgments

Many thanks to the children who participated in these tests and their parents, M.-J. Gagnon, J. Rioux and the École Du Touret, B. Côté and S.P.E.C. Tintamarre inc. for their collaboration.

Thanks also to the teams of engineering students involved in the design of the robots used in these tests. Research conducted by F. Michaud is supported by NSERC, CFI and FCAR.

References

[1] Kerstin Dautenhahn. Robots as social actors: Aurora and the case of autism. In *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco*, pages 359–374, 1999.

[2] Kerstin Dautenhahn. Socially intelligent agents and the primate social brain – towards a science of social minds. In *Technical Report FS-00-04, AAAI Fall Symposium on Socially Intelligent Agents - The Human in the Loop*, pages 35–51, 2000.

[3] F. Michaud and S. Caron. Roball – an autonomous toy-rolling robot. In *Proceedings of the Workshop on Interactive Robotics and Entertainment*, 2000.

[4] F. Michaud, A. Clavet, G. Lachiver, and M. Lucas. Designing toy robots to help autistic children - an open design project for electrical and computer engineering education. In *Proc. American Society for Engineering Education*, June 2000, 2000.

[5] Iain Werry and Kerstin Dautenhahn. Applying robot technology to the rehabilitation of autistic children. In *Proc. SIRS99, 7th International Symposium on Intelligent Robotic Systems '99*, 1999.

Chapter 16

AFFECTIVE SOCIAL QUEST

Emotion Recognition Therapy for Autistic Children

Katharine Blocher and Rosalind W. Picard
MIT Media Laboratory

Abstract This chapter describes an interactive computer system – Affective Social Quest – aimed at helping autistic children learn how to recognize emotional expressions. The system illustrates emotional situations and then cues the child to select which stuffed "dwarf" doll most closely matches the portrayed emotion. The dwarfs provide a wireless, playful haptic interface that can also be used by multiple players. The chapter summarizes the system design, discusses its use in behavioral modification intervention therapy, and presents evaluations of its use by six children and their practitioners.

1. Introduction

Recognizing and expressing affect is a vital part of social participation. Unfortunately, those with autism have a learning disability in this area, often accompanied by deficits in language, motor and perceptual development. Their development of social communication is very low compared to neurologically typical children who learn social cues naturally while growing up. In trying to comprehend social nuances in communication or social behavior to blend in during everyday interaction, autistic children get frustrated, not only with themselves but also with their teachers, and often give up learning. What may help an autistic child in this case is an ever-patient teacher. This research presents an approach to creating that teacher: a persistent and unresentful aid that progressively introduces basic emotional expressions, guides recognition development through matching, and records the child's success. It is designed to teach emotion recognition to autistic children with a heterogeneous disorder. Although the application developed for this research does not come close to the abilities of a highly trained human practitioner, it is designed to offload some of the more tedious parts of the work.

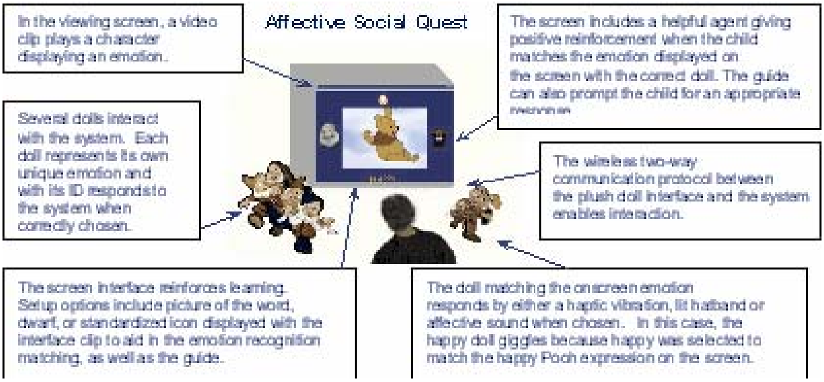
Affective Social Quest (ASQ) (figure 16.1) consists of a computer, custom software, and toy-like objects through which the child communicates to the computer. The system synthesizes interactive social situations in order to promote the recognition of affective information. This system will not tire because of impatience and can be a safe place for the child to explore. The goal of ASQ is to provide an engaging environment to help children – specifically autistic children – learn to recognize social displays of affect.

ASQ is an example of affective computing, research aimed at giving computers skills of emotional intelligence, including the ability to recognize and respond intelligently to emotion [3]. A computer can be taught to recognize aspects of emotion expression, such as facial movements indicative of a smile, and can prompt people for information related to human emotional state. However, computers are limited in their ability to recognize naturally occurring emotions; they can not easily generalize patterns from one situation to the next, nor do they understand the emotional significance associated with emotion expression. We recognize that some of the problems we face in trying to give computers emotion recognition abilities are similar to those therapists face in trying to help autistic children. We expect that progress in either of these areas will help inform progress in the other.

Six emotions that show up universally with characteristic facial expressions are: happiness, sadness, anger, fear, surprise, and disgust [2]. ASQ uses four of these: happiness, sadness, anger, and surprise, potentially displaying the emotion word, icon, doll face and representative video clips. The aim is to offer the child multiple representations for an emotion, to help him or her generalize many ways that one emotion may be displayed.

Different approaches for behavior intervention are available for autistic children. Many programs use emotion words and icon representations, showing children photographs of people exhibiting emotional expressions. However, systematic observations or experimental investigations of specific social behaviors are few ([1], [5], [4]). Many children with autism are drawn to computers, and can become engaged with off-the-shelf software. Most software applications for autistics focus on verbal development, object matching, or event sequencing. Laurette software is designed for autistic children to solve ‘what if’ scenarios and help them decide what the next action in a sequence could be. Mayer-Johnson has a "board maker" software tool that combines words with its standardized icons (Picture Communication Symbols (PCS)), to help children communicate through pictures (http://www.mayerjohnson.com/).

The ASQ system builds on the strengths of autistic children’s visual systems through use of video. Additionally, it incorporates characteristics of the intervention methods listed earlier. The potential for using affective computing and physical interfaces in therapy forms the heart of this work.



In the viewing screen, a video clip plays a character displaying an emotion.

The screen includes a helpful agent giving positive reinforcement when the child matches the emotion displayed on the screen with the correct doll. The guide can also prompt the child for an appropriate response

Several dolls interact with the system. Each doll represents its own unique emotion and with its ID responds to the system when correctly chosen.

The wireless two-way communication protocol between the plush doll interface and the system enables interaction.

The screen interface reinforces learning. Setup options include picture of the word, dwarf, or standardized icon displayed with the interface clip to aid in the emotion recognition matching, as well as the guide.

The doll matching the onscreen emotion responds by either a haptic vibration, lit hatband or affective sound when chosen. In this case, the happy doll giggles because happy was selected to match the happy Pooh expression on the screen.

Figure 16.1. Elements of the Interface.

2. The System

ASQ displays an animated show and offers pedagogical picture cues – the face of the plush dwarf doll, the emotion word, and the Mayer-Johnson standard icon – as well as an online guide that provides audio prompts to encourage appropriate response behavior from the child. The task was to have the system act as an ever-patient teacher. This led to a design focused on modeling antecedent interventions used in operant behavior conditioning. In essence, ASQ represents an automated discrete trial intervention tool used in behavior modification for teaching emotion recognition.

The system has multiple possibilities for interaction. In the default case, the system starts with a video clip displaying a scene with a primary emotion (antecedent) for the child to identify and match with the appropriate doll (target behavior). After a short clip plays, ASQ returns to a location in the clip and freezes on the image frame that reinforces the emotion that the child is prompted to select. The child is then prompted to indicate which emotion she recognizes in the clip, or frame - i.e., to select the appropriate doll matching that expression. To motivate interaction, the doll interface – the only input device to the system – creates a playful interaction for the child.

The practitioner can, using various windows, customize each interaction for each child. First, the practitioner can choose which video clips the child will be shown. These clips are arranged based on content (e.g. Cinderella), source (e.g. Animation), complexity (low-med-high difficulty of recognizing the emotion), duration (clip length), and emotion (happy, angry, sad, surprised). Practitioners may wish to focus on only a couple emotions early on, or may wish to avoid certain types of content depending on the idiosyncrasies of a particular child. The child's interface screen can be configured to include one or all of the following picture aids: Mayer-Johnson standardized icons representing the emotion, the word for that emotion, and a picture of the doll's (emotional) face. Also, an optional online animated guide can be included on the screen; it can be configured to provide an audible prompt (discriminative stimuli) for child interaction, or verbal reinforcement (putative response) for child affirmation.

The practitioner can configure many kinds of cues for ASQ to use in aiding the child. The dolls can cue the child with one of the following three choices: affect sound, hatband and lights, or internal vibration. The system can be cued to audibly play one of three sequences to prompt the child to select a doll to match the emotion in the video clip: for instance, when a happy clip plays, the system can say, "MATCH HAPPY" or "PUT WITH SAME", or "TOUCH HAPPY." Likewise, reinforcements for incorrect doll selections have three choices, such as "THAT'S SAD, MATCH HAPPY," etc. Seven different cue set-ups are configurable for one session with the timing, sequence, and repeat rate tailored for each. Additionally, the practitioner may opt to have the system play an entertaining reinforcement video clip, such as a Tigger song. Our objective was to offer as much flexibility to the practitioner as possible for customizing the screen interface for a particular session or specific child. This is especially important because autistic children often have unique idiosyncratic behaviors.

The child interface consists of one or more elements set up by the practitioner as just discussed. Figure 16.1 shows the screen seen by the child, set up here to show the video clip in the middle, the emotion icon at top, the dwarf face at left, the label of the emotion at bottom, and the guide at right. When selected, these images always appear in the same spot.

The child interacts with the system through a plush toy interface. Four interactive dwarves provide a tangible interface to the system, so that the child does not have to use a keyboard or mouse. Images of the dwarves, representing angry, happy, surprise, and sad, are pictured in figure 16.2, just as they appear on the screen when selected for display. The dolls serve as engaging input devices to the system: they are fun to hold and add a playful flavor to the interaction.

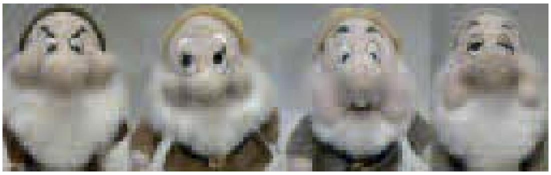


Figure 16.2. Pictures of the Dwarves.

The system design has two modes of interaction – an applied behavior mode and a story-based mode. The first mode displays short clips, one at a time, from various child program sources and the second mode displays an entire movie with the story segmented by the emotions. When the video freezes, the interaction is the same for both modes until the correct doll is selected. Working with researchers at the Dan Marino Center, Ft. Lauderdale, Florida, we designed the system to track performance information requested by the therapists. For each session, it recorded the child profiles, system configuration, clip configuration, and child response times[1].

3. Evaluation and Results

A pilot study was conducted to determine whether ASQ was engaging to children with autism and whether this type of application may potentially help children learn emotion recognition.

Subjects were recruited as volunteers through advertisements posted at the Dan Marino Center. Standardized assessment tools, as well as direct observation by trained psychologists and neurologists, were used to identify children whose primary deficits are related to social-emotional responding and appropriate affect. To participate in the pilot study, children needed to come to the center to play with ASQ for at least three days of sessions, each day's session lasting up to one hour. Nineteen children with deficits along the pervasive development disorder (PDD) or autism spectrum were exposed to ASQ. Six of these nineteen children were observed over three days. The therapy room was eight by eight feet, with one outside window and one window to another office. A laptop ran the ASQ application. The four dwarf dolls were the child's input devices to the application. Each toy doll was loosely positioned on the table on a reclining board adhered to the table with Velcro pads. The dolls could be picked up easily by the child, but were intended to remain on their stand because it was found to be easier for the child to press the belt-buckle of the chosen doll when the doll was on a hard surface (figure 16.3).

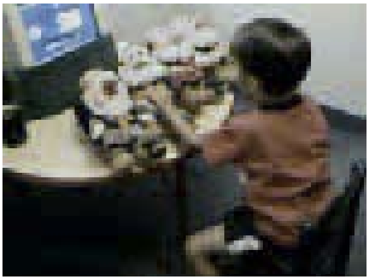


Figure 16.3. Child Testing.

The goal was to see if children can correctly match the emotion presented on the child-screen to the emotion represented by each doll. For experimental control the same dolls were used with each child, and all children were tested with the applied behavior mode (vs. story mode). The automated training was arranged to teach children to "match" four different emotion expressions: happy, sad, angry, and surprised. A standard discrete-trial training procedure with the automated application was used. Subjects sat facing the child-screen that exhibited specific emotional expressions under appropriate contexts within the child's immediate visual field. A video clip played for between 1 and 30 seconds. The clip displayed a scene in which an emotion was expressed by a character on the screen. The screen 'froze' on the emotional expression and waited for the child to touch the doll with the matching emotional expression (correct doll). After a pre-set time elapsed, the practitioner-cued sequence of visual and auditory prompts would be displayed.

If the child touched the doll with the corresponding emotional expression (correct doll), then the system affirmed the choice, e.g. the guide stated "Good, That's <correct emotion selected>," and an optional playful clip started to play on the child-screen. The application then displayed another clip depicting emotional content randomly pulled from the application.

If the child did not select a doll or if he selected the incorrect (non-matching) doll, the system would prompt, e.g. the guide would say "Match <correct emotion>" for no doll selection, or "That's <incorrect emotion>, Match <correct emotion>" for incorrect doll selection. The system waited for a set time configured by the practitioner and repeated its prompts until the child selected the correct doll. An optional replay of the clip could be set up before the session, in which case the application replays that same clip and proceeds with the specified order of prompts configured in the set up. If the child still fails to select the correct doll, the practitioner assists the child and repeats the verbal prompt and provides a physical prompt, e.g., pointing to the correct doll. If the child selects the correct doll but doesn't touch the doll after the physical prompt is provided, then physical assistance is given to insure that the child touches the correct doll. This procedure was used for the discrete trials.

Two low functioning autistic children, between the ages of 2 and 3, engaged in the video clips yet displayed little interest in the doll interface without direct assistance. One boy, age 4, demonstrated an understanding of the interaction, but struggled to match the appropriate doll. Another boy, aged 5, appeared to understand the interaction, yet had such a soft touch that he required assistance in touching the doll so that the system could detect what was selected.

A three-year-old child, with Spanish as native tongue, appeared very interested in the application regardless of the language difference. He and his family were visiting the US and played with ASQ for one hour. Earlier two visiting neurologists from Argentina sat in on the child session and they were certain that the screen interface had too many images (referring to the icon, word, and dwarf's face) and thought that the dolls were not a good interface. After they saw this boy interact with the application, both the physicians and the boy's parents were surprised at this boy's quick adaptation to the doll interface and his ability to recognize the emotions.

As suspected, higher functioning and older children, age 6-9, demonstrated ease with understanding the doll interaction, exhibited pleasure with the gaming aspect, and needed few of the helpful screen cues to make their selection. They were able to match emotional expressions displayed on their screen by selecting the correct doll after only a few sessions. One boy mimicked the displayed emotions on the screen. His mother reported that he was able to recognize other people's emotional expressions at home also.

We observed in some sessions that the child and a parent would each hold a doll, and when the parent was holding the doll that the child needed to select, the child would turn to the parent to achieve the selection; thus, the physical nature of the interface actually aided in helping the child with eye contact and shared experiences referred to as joint attention, another area where autistics often need help. Thus, we continue to see promise in the playful doll interface, despite occasional reliability problems with their sensors.

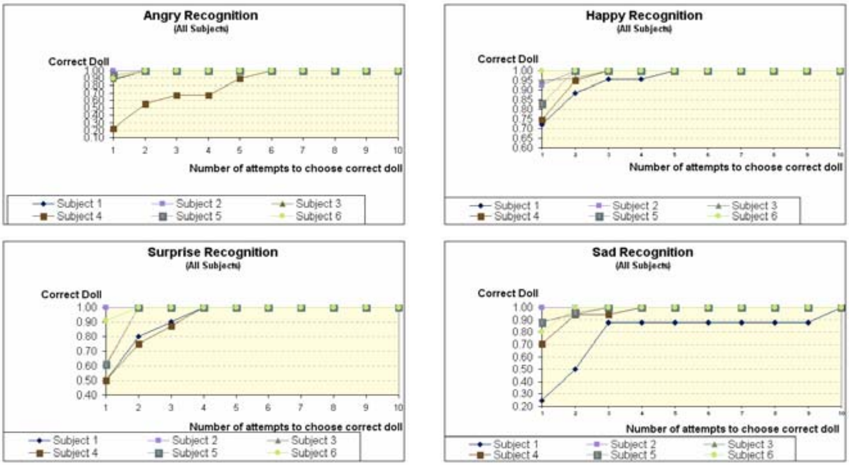


Figure 16.4. Recognition results for six kids (note some of the vertical scales differ.)

4. Conclusions

ASQ was successful at engaging the children. Furthermore, the statistical findings suggest that emotion matching occurred in most cases, with some children showing improvements in their performance over three sessions. Figure 16.4 shows combined recognition results for the six kids - where higher curves indicate better recognition rates. For example, when an angry clip was played, all kids (except subject 4) were able to correctly pick the angry doll when given two attempts. Subject 4 took many more tries than necessary to select the angry doll. For most of the kids, anger was the emotion most reliably recognized, while surprise and sadness were harder to get right. Five of the six kids were able to match many of the emotions on the first day. A three-year-old child showed results that he could recognize more samples of an emotion with each additional session of interaction. What the data did not provide is conclusive evidence that ASQ taught emotion recognition: it is possible that the children's performance improvement was due to something besides emotion recognition. A study including base line tests before and after using the system over a longer duration would present results that are more conclusive.

Although the ASQ system can measure improvements by a child while using the system, it does not assess improvements the child may show outside the computer world. One mother reported that her son said, "I'm happy" with a smile on his face at the dinner table with the family. She doesn't remember him expressing himself like that before. Also, she said that when he was picked up from school he asked if he could go play with the dwarves. Such feedback is promising; it needs to be gathered in a long-term systematic way in order to understand how the effects of the system generalize to real life.

Acknowledgments

We gratefully acknowledge Roberto Tuchman and David Lubin at the Dan Marino Center in Miami, Florida for their terrific help in evaluating the ASQ system.

Notes

1. We also computed: Response Rate to track the number of training trials (this measure includes both correct and incorrect responses by the child, normalized by the time of the session), Accuracy as an index of how effective the training procedures are for teaching the children to match the targeted emotion (consists of a ratio of correct matches over total attempted matches, for each trial), and Fluency as a performance summary of how many correct responses were made (this measure combines response rate and accuracy). An accompanying thesis [1] provides formulas for these measures.

References

[1] K. Blocher. Affective social quest: Teaching emotion recognition with interactive media and wireless expressive toys. *SM Thesis, MIT*, 1999.

[2] P. Eckman. An argument for basic emotions. *Cognition and Emotion*, 6, 1992.

[3] R. Picard. *Affective Computing*. MIT Press, Cambridge, MA, 1997.

[4] M. Sigman and L. Capps. *Children with Autism : A Developmental Perspective*. Harvard Press, Cambridge, MA, 1997.

[5] R. Tuchman. Personal communication. Conversation while conducting experiments at Dan Marino Center in September 1999, 1999.

Chapter 17

PEDAGOGICAL SOAP

Socially Intelligent Agents for Interactive Drama

Stacy C. Marsella
USC Information Sciences Institute

Abstract Interactive Pedagogical Dramas (IPD) are compelling stories that have didactic purpose. Autonomous agents realize the characters in these dramas, with roles that require them to interact with a depth and subtlety consistent with human behavior in difficult, stressful situations. To address this challenge, the agent design is based on psychological research on human emotion and behavior. We discuss our first IPD, *Carmen's Bright IDEAS*, an interactive drama designed to improve the social problem solving skills of mothers of pediatric cancer patients.

1. Introduction

Carmen is the mother of a seriously ill nine-year-old boy, Jimmy. Jimmy's illness is a significant physical and emotional drain on Carmen and her family. Carmen is often at the hospital with Jimmy. As a result, Carmen's six-year-old daughter, Diana, is having temper tantrums because she feels scared and neglected. Carmen's boss is also upset about her absences from work. Unable to effectively deal with these problems, Carmen is experiencing high levels of psychological distress, including anxiety and depression. To help her address these problems, a clinical counselor, Gina, is going to train Carmen in a problem-solving technique called Bright IDEAS.

The above is the background story of *Carmen's Bright IDEAS*, an *interactive pedagogical drama* (IPD) realized by socially intelligent agents. *Carmen's Bright IDEAS* is designed to improve the problem solving skills of mothers of pediatric patients, mothers that face difficulties similar to Carmen's. The pedagogical goal of the drama is to teach a specific approach to social decision-making and problem solving called Bright IDEAS. Each letter of IDEAS refers to a separate step in the problem solving method (*Identify* a solvable problem, *Develop* possible solutions, *Evaluate* your options, *Act* on your plan and *See* if it worked).

In an interactive pedagogical drama, a learner (human user) interacts with believable characters in a believable story that the learner empathizes with. In particular, the characters may be facing and resolving overwhelming, emotionally charged difficulties *similar* to the learner's. The learner's identification with the characters and the believability of their problems are central to the goals of having the learner fully interact with the drama, accept the efficacy of the skills being employed in it and subsequently apply those skills in her own life.

The design of IPDs poses many challenges. The improvisational agents who answer the casting call for characters like Carmen and Gina must provide convincing portrayals of humans facing difficult personal and social problems. They must have ways of modeling goals, personality and emotion, as well as ways of portraying those models via communicative and evocative gestures.

Most critically, an IPD is a social drama. Thus, the agents in the drama must behave like socially interacting humans. An agent has to be concerned with how other agents view their behavior. They may emotionally react if they believe others view them in an way that is inconsistent with how they see themselves (their ego identity). Also, to achieve its goals, an agent may need to motivate, or manipulate, another agent to act (or not to act).

Due to the highly emotional, stressful events being dramatized, the design of the agent models was a key concern. The design was heavily inspired by emotional and personality models coming out of work on human stress and coping (Lazarus 1991), in contrast to the more commonly used models in agent design coming out of a cognitive or linguistic view (e.g., [6], [10], [11]).

IPDs are animated dramas and therefore their design raises a wide range of presentational issues and draws on a range of research to address those issues that can only be briefly touched upon here (see [8] for additional details). The agent architecture uses a model of gesture heavily influenced not only by work on communicative use of gesture ([3], [9]) but also work on non-communicative but emotionally revealing nonverbal behavior [4], including work coming out of clinical studies [5]. Further, since these agents are acting out in a drama, there must be ways to dynamically manage the drama's structure and impact even while the characters in it are self-motivated, improvisational agents (e.g., [7], [2]). Because IPDs are animated and dynamically unfold, there must be ways of managing their presentation (e.g., [1], [12]).

The discussion that follows provides a brief overview of the IPD design. The relation of the agents' emotional modeling to their social interactions is then discussed in greater detail using examples drawn from Carmen's Bright IDEAS.

2. IPD Background

In our basic design for interactive pedagogical drama, there are five main components: a cast of autonomous character agents, the 2D or 3D puppets which are the physical manifestations of those agents, a director agent, a cinematographer agent, and finally the learner/user who impacts the behavior of the characters. Animated agents in the drama choose their actions autonomously following directions from the learner and/or a director agent. Director and cinematographer agents manage the interactive drama's onscreen action and its presentation, respectively, so as to maintain story structure, achieve pedagogical goals, and present the dynamic story so as to achieve best dramatic effect. The design of all these agents requires both general capabilities as well as knowledge specific to the interactive drama that is being created.

Our current approach to the design of IPDs is to start with a professionally written script and systematically deconstruct it. The deconstruction serves several ends. It provides a model of the story and how variability can enter that story. In particular, the deconstruction provides the knowledge to dynamically direct the agents in the drama. It also guides the modeling of the improvisational agents in the drama, their personalities, their goals, their dialog, as well as how they interact to achieve their goals. Finally, it serves to constrain the complexity of these models. Detailed discussion of this script deconstruction approach and the overall IPD architecture is beyond the scope of this chapter but more details can be found in [8].

2.1 Carmen's Bright IDEAS

The story for Carmen's Bright IDEAS is organized into three acts. The first act reveals the back story. The second, main, act takes place in Gina's office. Carmen discusses her problems with Gina, who suggests she use Bright IDEAS to help her find solutions. See Figure 17.1. With Gina's help, Carmen goes through the initial steps of Bright IDEAS, applying the steps to one of her problems and then completes the remaining steps on her own. The final act reveals the outcomes of Carmen's application of Bright IDEAS to her problems.

The human mother interacts with the drama by making choices for Carmen such as what problem to work on, what Carmen's inner thoughts are at critical junctures, etc. The mother's selection of inner thoughts for Carmen impacts her emotional state, which in turn impacts her thoughts and behavior. It is Gina's task to keep the social problem solving on track by effectively responding to Carmen's state, and motivating her through dialog. Meanwhile, a bodiless cinematographer agent is dynamically manipulating the camera views, flashbacks, and flash-forwards.

Figure 17.1. Carmen in Gina's office.

Gina and Carmen interact through spoken dialog. In order to add realism and maximize the expressive effect of this dialog, recorded dialog of voice actors is used instead of speech synthesis. A significant amount of variability in the generated dialog is supported by breaking the recordings into meaningful individual phrases and fragments. Additionally variability is achieved by recording multiple variations of the dialog (in content and emotional expression). The agents compose their dialog on the fly. The dialog is also annotated with its meaning, intent and emotional content. The agents use the annotations to understand each other, to decide what to say, and more generally to interact. The agents experience the annotations in order, so their internal state and appearance can be in flux over the dialog segment.

3. Agent Architecture

The agent architecture is depicted in Figure 17.2. There are modules for *problem solving*, *dialog*, *emotional appraisal* and *physical focus*. The problem solving module is the agent's cognitive layer, specifically its goals, planning and deliberative reaction to world events. The dialog module models how to use dialog to achieve goals. Emotional appraisal is how the agent emotionally evaluates events (e.g., the dialog annotations). Finally, physical focus manages the agent's nonverbal behavior.

There are several novel pathways in the model worth noting. The agent's own acts feed back as input. Thus it is possible for the agent to say something and then emotionally and cognitively react to the fact that it has said it. Emotional appraisal impacts problem solving, dialog and behavior. Finally, there are multiple inputs to physical focus, from emotional appraisal, dialog and problem solving, all competing for the agent's physical resources (arms, legs, mouth, head, etc.). For instance, the dialog module derives dialog that

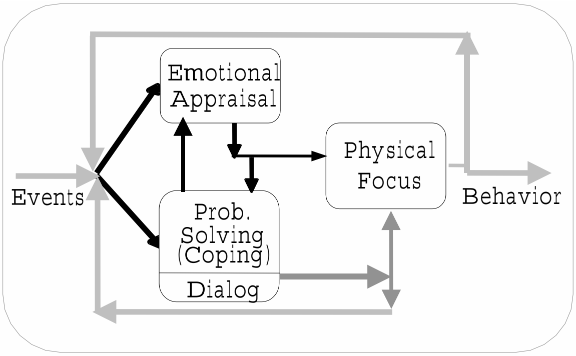


Figure 17.2. Agent Architecture.

it intends to communicate, which may include an intent to project an associated emotion. This communication may be suggestive of certain nonverbal behavior for the agent's face, arms, hands etc. However, the agent's emotional state derived from emotional appraisal may suggest quite different behaviors. Physical focus mediates this contention.

A simple example demonstrates how some of these pathways work. Gina may ask Carmen why her daughter is having temper tantrums. Feeling anxious about being judged a bad mother, Carmen copes (problem solving) by dismissing the significance of the tantrums (dialog model): "She is just being babyish, she wants attention." Based on Carmen's dialog and emotional state, physical focus selects relevant behaviors (e.g., fidgeting with her hands). Her dialog also feeds back to emotional appraisal. She may now feel guilty for "de-humanizing" her child, may physically display that feeling (physical focus) and then go on to openly blame herself. Carmen can go through this sequence of interactions solely based on the flux in her emotional reaction to her own behavior. Gina, meanwhile, will emotionally appraise Carmen's seeming callousness and briefly reveal shock (e.g., by raised eyebrows), but that behavior may quickly be overridden if her dialog model decides to project sympathy.

Emotional appraisal plays a key role in shaping how the agents interact and how the user interacts with Carmen. The appraisal model draws on the research of Richard Lazarus (1991). In the Lazarus model, emotions flow out of cognitive appraisal and management of the person-environment relationship. Appraisal of events in terms of their significance to the individual leads to emotions and tendencies to cope in certain ways. The appraisal process is broken into two classes. Primary appraisal establishes an event's relevance. Secondary appraisal addresses the options available to the agent for coping with the event. One of the key steps in primary appraisal is to determine an individual's ego involvement: how an event impacts the agent's collection of individual commitments, goals, concerns or values that comprise its ego-identity. This models concerns for self and social esteem, social roles, moral values, concern for other people and their well-being and ego-ideals. In IPD, the knowledge modeled by the agent's ego identity comprises a key element of how it interacts with other characters and its response to events. For example, it is Carmen's concern for her son's well-being that induces sadness. And it is her ideal of being a good mother, and desire to be perceived as one (social esteem), that leads to anxiety about discussing Diana's tantrums with Gina.

The emotional appraisal module works with the dialog module to create the rich social interactions necessary for dramas like Carmen's Bright IDEAS. Dialog socially obligates the listening agent to respond and may impact their emotional state, based on their emotional appraisal. The IPD dialog module currently models several dialog moves; Suggest (e.g., an approach to a problem), Ask/Prompt (e.g., for an answer), Re-Ask/Re-Prompt, Answer, Reassure (e.g., to impact listener's emotional state), Agree/Sympathize (convey sympathy), Praise, Offer-Answer (without being asked), Clarify (elaborate) and Resign (give-up). The agent chooses between these moves depending on dialog state as well as the listener's emotional state. In addition, an intent to convey emotional state, perhaps distinct from the agent's appraisal-based emotional state, is derived from these moves.

3.1 Interactions from 3 Perspectives

To exemplify how the agents socially interact, it is useful to view it from multiple perspectives. From Gina's perspective, the social interaction is centered around a persistent goal to motivate Carmen to apply the steps of the IDEAS approach to her problems. This goal is part of the knowledge stored in Gina's problem solving module (and is also part of her ego identity). Dialog is Gina's main tool in this struggle and she employs a variety of dialog strategies and individual dialog moves to motivate Carmen. An example of a strategy is that she may ask Carmen a series of questions about her problems that will help Carmen identify the causes of the problems. At a finer-grain, a variety of dialog moves may be used to realize the steps of this strategy. Gina may reassure Carmen that this will help her, prompt her for information or praise her. Gina selects between these moves based on the dialog state and Carmen's emotional state. The tactics work because Gina's dialog (the annotations) will impact Carmen emotionally and via obligations.

Carmen has a different perspective on the interaction. Carmen is far more involved emotionally. The dialog with Gina is a potential source of distress, due to the knowledge encoded in her emotional appraisal module. For example, her ego involvement models concern for her children, desire to be viewed as a good mother as well as inference rules such as "good mothers can control their children" and "treat them with respect." So discussing her daughter's tantrums can lead to sadness out of concern for Diana and anxiety/guilt because failure to control Diana may reflect on her ability as a mother. More generally, because of her depression, the Carmen agent may initially require prompting. But as she is reassured, or the various subproblems in the strategy are addressed, she will begin to feel hopeful that the problem solving will work and may engage the problem solving without explicit prompting.

The learner is also part of this interaction. She impacts Carmen by choosing among possible thoughts and feelings that Carmen might have in the current situation, which are then incorporated into Carmen's mental model, causing Carmen to act accordingly. This design allows the learner to adopt different relationships to Carmen and the story. The learner may have Carmen feel as she would, act they way she would or "act out" in ways she would not in front of her real-world counselor.

The combination of Gina's motivation through dialog and the learner's impact on Carmen has an interesting impact on the drama. While Gina is using dialog to motivate Carmen, the learner's interaction is also influencing Carmen's thoughts and emotions. This creates a tension in the drama, a tug-of-war between Gina's attempts to motivate Carmen and the initial, possibly less positive, attitudes of the Carmen/learner pair. As the learner plays a role in determining Carmen's attitudes, she assumes a relationship in this tug-of-war, including, ideally, an empathy for Carmen and her difficulties, a responsibility for the onscreen action and perhaps empathy for Gina. If Gina gets Carmen to actively engage in applying the IDEAS technique with a positive attitude, then she potentially wins over the learner, giving her a positive attitude. Regardless, the learner gets a vivid demonstration of how to apply the technique.

4. Concluding Comments

The social interactions in *Carmen's Bright IDEAS* are played out in front of a demanding audience - mothers undergoing problems similar to Carmen. This challenges the agents to socially interact with a depth and subtlety consistent with human behavior in difficult, stressful situations. Currently, the *Carmen's Bright IDEAS* prototype is in clinical trials, where it is facing its demanding audience. The anecdotal feedback is extremely positively. Soon, a careful evaluation of how well the challenge has been addressed will be forthcoming.

Acknowledgments

The work discussed here was done with W. Lewis Johnson and Catherine LaBore. The author also wishes to thank our clinical psychologist collaborators, particularly O.J. Sahler, MD, Ernest Katz, Ph.D., James Varni, Ph.D., and Karin Hart, Psy.D. Discussions with Jeff

Rickel and Jon Gratch were invaluable. This work was funded by the National Cancer Institute under grant R25 CA65520-04.

References

[1] W.H. Bares and J. C. Lester. Intelligent multi-shot visualization interfaces for dynamic 3d worlds. In M. Maybury, editor, *Proc. International Conference on Intelligent User Interfaces, Redondo Beach, CA*, pages 119–126. ACM Press, 1999.

[2] B. Blumberg and T. Galyean. Multi-level direction of autonomous creatures for real-time virtual environments. In *Computer Graphics (SIGGRAPH 95 Proceedings)*, pages 47–54. ACM SIGGRAPH, 1995.

[3] J. Cassell and M. Stone. Living hand to mouth: Psychological theories about speech and gesture in interactive dialogue systems. *Psychological Models of Communication in Collaborative Systems, AAAI Fall Symposium 1999*, AAAI Press, pp. 34-42, 1999.

[4] P. Ekman and W. V. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, 1:49–97, 1969.

[5] N. Freedman. The analysis of movement behavior during clinical interview. In A. Siegman and B. Pope, editors, *Studies in Dyadic Communication*, pages 153–175. New York: Pergamon Press, 1997.

[6] N. Frijda. *The emotions*. Cambridge University Press, 1986.

[7] M. T. Kelso, P. Weyhrauch, and J. Bates. Dramatic presence. *Presence: Journal of Teleoperators and Virtual Environments*, 2(1), 1993.

[8] S. C. Marsella, W. L. Johnson, and C. LaBore. Interactive pedagogical drama. In C. Sierra, M. Gini, and J. S. Rosenschein, editors, *Proc. Fourth International Conference on Autonomous Agents, Barcelona, Spain*, pages 301–308. ACM Press, 2000.

[9] D. McNeil. *Hand and Mind*. University of Chicago Press, Chicago, 1992.

[10] D. Moffat. Personality parameters and programs. In R. Trappl and P. Petta, editors, *Creating Personalities for Synthetic Actors*, pages 120–165. Springer, 1997.

[11] K. Oatley and P.N. Johnson-Laird. Towards a cognitive theory of emotions. *Cognition and Emotion*, 1(1):29–50, 1987.

[12] B. Tomlinson, B. Blumberg, and D. Nain. Expressive autonomous cinematography for interactive virtual environments. In C. Sierra, M. Gini, and J. S. Rosenschein, editors, *Proc. Fourth International Conference on Autonomous Agents, Barcelona, Spain*, pages 317–324. ACM Press, 2000.

Chapter 18

DESIGNING SOCIABLE MACHINES

Lessons Learned

Cynthia Breazeal
MIT Media Lab

Abstract Sociable machines are a blend of art, science, and engineering. We highlight how insights from these disciplines have helped us to address a few key design issues for building expressive humanoid robots that interact with people in a social manner.

1. Introduction

What is a sociable machine? In our vision, a sociable machine is able to communicate and interact with us, understand and even relate to us, in a personal way. It should be able to understand us and itself in social terms. We, in turn, should be able to understand it in the same social terms—to be able to relate to it and to empathize with it. In short, a sociable machine is socially intelligent in a human-like way, and interacting with it is like interacting with another person [7].

Humans, however, are the most socially advanced of all species. As one might imagine, an autonomous humanoid robot that could interpret, respond, and deliver human-style social cues even at the level of a human infant is quite a sophisticated machine. For the past few years, we have been exploring the simplest kind of human-style social interaction and learning (that which occurs between a human infant with its caregiver) and have used this as a metaphor for building a sociable robot, called Kismet. This is a scientific endeavor, an engineering challenge, and an artistic pursuit. This chapter discusses a set of four design issues underlying Kismet's compelling, life-like behavior, and the lessons we have learned in building a robot like Kismet.

2. Designing Sociable Robots

Somewhat like human infants, sociable robots shall be situated in a very complex social environment (that of adult humans) with limited perceptual, motor, and cognitive abilities. Human infants, however, are born with a set of perceptual and behavioral biases. Soon after birth they are particularly attentive to people and human-mediated events, and can react in a recognizable manner (called proto-social responses) that conveys social responsiveness. These innate abilities suggests how critically important it is for the infant to establish a social bond with his caregiver, both for survival purposes as well as to ensure normal cognitive and social development [4]. For this reason, Kismet has been given a roughly analogous set of perceptual and behavioral abilities (see Figure 18.1, and refer to [3] for technical details).

Together, the infant's biological attraction to human-mediated events in conjunction with his proto-social responses launch him into social interactions with his caregiver. There is an imbalance in the social and cultural sophistication of the two partners. Each, however, has innate endowments for helping the infant deal with a rich social environment. For instance, the infant uses protective responses and expressive displays for avoiding harmful or unpleasant situations and to encourage and engage in beneficial ones. Human adults seem to intuitively read these cues to keep the infant comfortable, and to adjust their own behavior to suit his limited perceptual, cognitive, and motor abilities.

Being situated in this environment is critical for normal development because as the infant's capabilities improve and become more diverse, there is still an environment of sufficient complexity into which he can develop. For this reason, Kismet has been designed with mechanisms to help it cope with a complex social environment, to tune its responses to the human, and to give the human social cues so that she is better able to tune herself to it. This allows Kismet to be situated in the world of humans without being overwhelmed or under-stimulated.

Both the infant's responses and his parent's own caregiving responses have been selected for because they encourage adults to treat the infant as an intentional being—as if he is already fully socially aware and responsive with thoughts, wishes, intents, desires, and feelings that he is trying to communicate as would any other person. This "deception" is critical for the infant's development because it bootstraps him into a cultural world [4]. Over time, the infant discovers what sorts of activity on his part will get responses from her, and also allows for routine, predictable sequences to be established that provide a context of mutual expectations. This is possible due to the caregiver's consistent and predictable manner of responding to her infant *because* she assumes that he is fully socially responsive and shares the same meanings that she applies to the interaction. Eventually, the infant exploits these con-

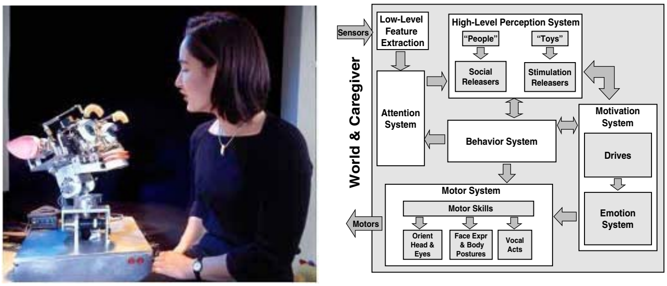


Figure 18.1. Kismet (left) has 15 degrees of freedom (DoF) in its face, 3 for the eyes, and 3 for the neck. It has 4 cameras, one behind each eyeball, one between the eyes, and one in the "nose." It can express itself through facial expression, body posture, gaze direction, and vocalizations. The robot's architecture (right) implements perception, attention, behavior arbitration, motivation (drives and emotive responses) and motor acts (expressive and skill oriented).

sistencies to learn the significance his actions and expressions have for other people so that he *does* share the same meanings. This is the sort of scenario that we are exploring with Kismet. Hence, it is important that humans treat and respond to Kismet in a similar manner, and Kismet has been designed to encourage this.

Regulation of Interactions. As with young infants, Kismet must be well-versed in regulating its interactions with the caregiver to avoid becoming overwhelmed or under-stimulated. Inspired by developmental psychology, Kismet has several mechanisms for accomplishing this, each for different kinds of interactions. They all serve to slow the human down to an interaction rate that is within the comfortable limits of Kismet's perceptual, mechanical, and behavioral limitations. Further, Kismet provides readable cues as to what the appropriate level of interaction is. The robot exhibits interest in its surroundings and in the humans that engage it, and behaves in a way to bring itself closer to desirable aspects and to shield itself from undesirable aspects. By doing so, Kismet behaves to promote an environment for which its capabilities are well-matched—ideally, an environment where it is slightly challenged but largely competent—in order to foster its social development.

We have found two distinct regulatory systems to be effective in helping Kismet to maintain itself in a state of "well-being." These are the emotive responses and the homeostatic regulatory mechanisms. The drive processes establish the desired stimulus and motivate the robot to seek it out and to engage it. The emotions are another set of mechanisms (see Table 18.1), with greater direct control over behavior and expression, that serve to bring the robot closer to desirable situations ("joy," "interest," even "sorrow"), and cause the robot to withdraw from or remove undesirable situations ("fear," "anger," or "disgust").

Which emotional response becomes active depends largely on the perceptual releasers, but also on the internal state of the robot. The behavioral strategy may involve a social cue to the caregiver (through facial expression and body posture) or a motor skill (such as the escape response). We have found that people readily read and respond to these expressive cues. The robot's use of facial displays to define a personal space is a good example of how social cues, that are a product of emotive responses, can be used to regulate the proximity of the human to the robot to benefit the robot's visual processing [3].

Table 18.1. Summary of the antecedents and behavioral responses that comprise Kismet's emotive responses. The antecedents refer to the eliciting perceptual conditions for each emotion process. The behavior column denotes the observable response that becomes active with the "emotion." For some, this is simply a facial expression. For others, it is a behavior such as escape. The column to the right describes the function each emotive response serves Kismet.

Antecedent Conditions	*Emotion*	*Behavior*	*Function*
Delay, difficulty in achieving goal of adaptive behavior	anger, frustration	complain	show displeasure to caregiver to modify his/her behavior
Presence of an undesired stimulus	disgust	withdraw	signal rejection of presented stimulus to caregiver
Presence of a threatening, overwhelming stimulus	fear, distress	escape	Move away from a potentially dangerous stimuli
Prolonged presence of a desired stimulus	calm	engage	Continued interaction with a desired stimulus
Success in achieving goal of active behavior, or praise	joy	display pleasure	Reallocate resources to the next relevant behavior (or reinforce behavior)
Prolonged absence of a desired stimulus, or prohibition	sorrow	display sorrow	Evoke sympathy and attention from caregiver (or discourage behavior)
A sudden, close stimulus	surprise	startle response	alert
Appearance of a desired stimulus	interest	orient	attend to new, salient object
Need of an absent and desired stimulus	boredom	seek	Explore environment for desired stimulus

Establishment of Appropriate Social Expectations. It will be quite a while before we are able to build autonomous humanoids that rival the social competence of human adults. For this reason, Kismet is designed to have an infant-like appearance of a fanciful robotic creature. Note that the human is a critical part of the environment, so evoking appropriate behaviors from the human is essential for this project. Kismet should have an appealing appearance and a natural interface that encourages humans to interact with Kismet as if it were a young, socially aware creature. If successful, humans will naturally and unconsciously provide scaffolding interactions. Furthermore, they will expect the robot to behave at a competency-level of an infant-like creature. This level should be commensurate with the robot's perceptual, mechanical, and computational limitations.

Great care has been taken in designing Kismet's physical appearance, its sensory apparatus, its mechanical specification, and its observable behavior (motor acts and vocal acts) to establish a robot-human relationship that adheres to the infant-caregiver metaphor. Following the baby-scheme of Eibl-Eiblsfeldt [8], Kismet's appearance encourages people to treat it as if it were a very young child or infant. Kismet has been given a child-like voice and it babbles in its own characteristic manner.

Given Kismet's youthful appearance, we have found that people use many of the same behaviors that are characteristic of interacting with infants. As a result, they present a simplified class of stimuli to the robot's sensors, which makes our perceptual task more manageable without having to explicitly instruct people in how to engage the robot. For instance, we have found that people intuitively slow down and exaggerate their behavior when playing with Kismet, which simplifies the robot's perceptual task. Female subjects are willing to use exaggerated prosody when talking to Kismet, characteristic of *motherese*. Both male and female subjects tend to sit directly in front of and close to Kismet, facing it the majority of the time. When engaging Kismet in proto-dialogue, they tend to slow down, use shorter phrases, and wait longer for Kismet's response. Some subjects use exaggerated facial expressions.

Along a similar vein, the design should minimize factors that could detract from a natural infant-caretaker interaction. Ironically, humans are particularly sensitive (in a negative way) to systems that try to imitate humans but inevitably fall short. Humans have strong implicit assumptions regarding the nature of human-like interactions, and they are disturbed when interacting with a system that violates these assumptions [6]. For this reason, we consciously decided to *not* make the robot look human.

Readable Social Cues. As with human infants, Kismet should send social signals to the human caregiver that provide the human with feedback of its internal state. This allows the human to better predict what the robot is likely to do and to shape their responses accordingly. Kismet does this by means of expressive behavior. It can communicate emotive state and social cues to a human through facial expression, body posture, gaze direction, and voice. We have found that the scientific basis for how emotion correlates to facial expression [12] or vocal expression [10, 5] to be very useful in mapping Kismet's emotive states to its face actuators and its articulatory-based speech synthesizer. Results from various forced-choice and similarity studies suggest that Kismet's emotive facial expressions and vocal expressions are readable.

Furthermore, we have learned that artistic insights complement these scientific findings in very important ways. A number of animation guidelines and techniques have been developed for achieving life-like, believable, and compelling animation [13, 11]. These rules of thumb are designed to create behavior that is rich and interesting, yet easily understandable to the human observer. For instance, animators take a lot of care in drawing the audience's attention to the right place at the right time. To enhance the readability and understandability of Kismet's behavior, Kismet's expression and gaze precede its behavioral response to make its behavior understandable and predictable to the human who interacts with it. People naturally tend to look at what Kismet is looking at. They observe the expression on its face to see how the robot will respond towards it. If the robot has a frightened expression, the observer is not surprised to witness a fleeing response soon afterwards. If they are behaving towards the robot in a way that generates a negative expression, they soon correct their behavior.

By incorporating these scientific and artistic insights, we found that people intuitively and naturally use Kismet's expressive feedback to tune their performance in the exchange. We have learned that through a process of entraining to the robot, both the human and robot benefit: the person enjoys the easy interaction while the robot is able to perform effectively within its perceptual, computational, and behavioral limits. Ultimately, these cues will allow humans to improve the quality of their instruction. For instance, human-robot entrainment can be observed during turn-taking interactions. They start to use shorter phrases, wait longer for the robot to respond, and more carefully watch the robot's turn-taking cues. The robot prompts the other for his/her turn by craning its neck forward, raising its brows, and looking at the person's face when it's ready for him/her to speak. It will hold this posture for a few seconds until the person responds. Often, within a second of this display, the subject does so. The robot then leans back to a neutral posture, assumes a neutral expression, and tends to shift its gaze away from the person. This cue indicates that the robot is about to speak. The robot typically issues one utterance, but it may issue several. Nonetheless, as the exchange proceeds, the subjects tend to wait until prompted. This allows for longer runs of clean turns before an interruption or delay occurs in the robot-human proto-dialogue.

Interpretation of Human's Social Cues. During social exchanges, the person sends social cues to Kismet to shape its behavior. Kismet must be able to perceive and respond to these cues appropriately. By doing so, the quality of the interaction improves. Furthermore, many of these social cues will eventually be offered in the context of teaching the robot. To be able to take advantage of this scaffolding, the robot must be able to correctly interpret and react to these social cues. There are two cases where the robot can read the human's social cues.

The first is the ability to recognize praise, prohibition, soothing, and attentional bids from robot-directed speech [9, 2]. This could serve as an important teaching cue for reinforcing and shaping the robot's behavior. Several interesting interactions have been witnessed between Kismet and human subjects when Kismet recognizes and expressively responds to their tone of voice. They use Kismet's facial expression and body posture to determine when Kismet "understood" their intent. The video of these interactions suggests evidence of affective feedback where the subject might issue an intent (say, an attentional bid), the robot responds expressively (perking its ears, leaning forward, and rounding its lips), and then the subject immediately responds in kind (perhaps by saying, "Oh!" or, "Ah!"). Several subjects appeared to empathize with the robot after issuing a prohibition—often reporting feeling guilty or bad for scolding the robot and making it "sad."

The second is the ability of humans to direct Kismet's attention using natural cues [1]. This could play an important role in socially situated learning by giving the caregiver a way of showing Kismet what is important for the task, and for establishing a shared reference. We have found that it is important for the robot's attention system to be tuned to the attention system of humans. It is important that both human and robot find the same types of stimuli salient in similar conditions. Kismet has a set of perceptual biases based on the human pre-attentive visual system. In this way, both robot and humans are more likely to find the same sorts of things interesting or attention-grabbing. As a result, people can very naturally and quickly direct the robot's attention by bringing the target close and in front of the robot's face, shaking the object of interest, or moving it slowly across the centerline of the robot's face. Each of these cues increases the saliency of a stimulus by making it appear larger in the visual field, or by supplementing the color or skin-tone cue with motion. Kismet's attention system coupled with gaze direction provides people with a powerful and intuitive social cue for when they have succeeded in steering the robot's interest.

3. Summary

In this chapter, we have outlined a set of four core design issues that have guided our work in building Kismet. When engaging another socially, humans bring a complex set of well-established social machinery to the interaction. Our aim is not a matter of re-engineering the human side of the equation to suit the robot. Instead, we want to engineer *for* the human side of the equation—to design Kismet in such a way to support what comes naturally to people, so that they will intuitively communicate with and teach the robot. Towards this, we have learned that both artistic and scientific insights play an important role in designing sociable robots that follow the infant-caregiver metaphor. The design encourages people to intuitively engage in appropriate interactions with the robot, from which we can explore socially situated learning scenarios.

Acknowledgments

The author gratefully acknowledges the creativity and ingenuity of the members of the Humanoid Robotics Group at the MIT Artificial Intelligence Lab. This work was funded by NTT and Darpa contract DABT 63–99–1–0012.

References

[1] C. Breazeal and B. Scassellati. A Context-Dependent Attention System for a Social Robot. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99)*, pages 1146–1151, Stockholm, Sweden, 1999.

[2] C. Breazeal and L. Aryananda. Recognition of Affective Communicative Intent in Robot-Directed Speech. In *Proceedings of the First IEEE-RAS International Conference on Humanoid Robots (Humanoids2000)*, Cambridge, MA, 2000.

[3] C. Breazeal. *Designing Sociable Robots*. MIT Press, Cambridge, MA, 2002.

[4] M. Bullowa, editor. *Before Speech: The Beginning of Interpersonal Communication*. Cambridge University Press, Cambridge, UK, 1979.

[5] J. Cahn. *Generating Expression in Synthesized Speech*. S.M. thesis, Massachusetts Institute of Technology, Department of Media Arts and Sciences, Cambridge, MA, 1990.

[6] J. Cole. *About Face*. MIT Press, Cambridge, MA, 1998.

[7] K. Dautenhahn. The Art of Designing Socially Intelligent Agents: Science, Fiction, and the Human in the Loop. in *Applied Artificial Intelligence*, 12(7–8): 573–617, 1998.

[8] I. Eibl-Eibesfeldt. Similarities and differences between cultures in expressive movements. In R. Hinde, editor, *Nonverbal Communication*, pages 297–311. Cambridge University Press, Cambridge, UK, 1972.

[9] A. Fernald. Intonation and communicative intent in mother's speech to infants: Is the melody the message? *Child Development*, 60: 1497–1510, 1989.

[10] I. Murray and L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal Acoustical Society of America*, 93(2): 1097–1108, 1993.

[11] F. Parke and K. Waters. *Computer Facial Animation*. A. K. Peters, Wellesley, MA, 1996.

[12] C. Smith and H. Scott. A Componential Approach to the Meaning of Facial Expressions. In J. Russell and J.M. Fernández-Dols, editors, *The Psychology of Facial Expression*, pages 229–254. Cambridge University Press, Cambridge, UK, 1997.

[13] F. Thomas and O. Johnston. *Disney Animation: The Illusion of Life*. Abbeville Press, New York, 1981.

Chapter 19

INFANOID

A Babybot that Explores the Social Environment

Hideki Kozima
Communications Research Laboratory

Abstract We are building an infant-like robot, *Infanoid*, to investigate the underlying mechanisms of *social intelligence* that will allow it to communicate with human beings and participate in human social activities. We propose an ontogenetic model of social intelligence, which is being implemented in *Infanoid*: how the robot acquires communicative behavior through interaction with the social environment, especially with human caregivers. The model has three stages: (1) the acquisition of *intentionality*, which enables the robot to make use of certain methods for obtaining goals, (2) *identification* with others, which enables it to indirectly experience others' behavior, and (3) *social communication*, in which the robot understands others' behavior by ascribing it the intention that best explains the behavior.

1. Introduction

Imagine a robot that can understand and produce a *complete* repertoire of human communicative behavior, such as gestures and language. However, when this robot encounters novel behavior, it fails to understand it. Or, if the robot encounters a novel situation where any behavior in its repertoire does not work at all, it gets stuck. As long as the robot is preprogrammed according to a blueprint, it is best to take a *design stance*, instead of a *intentional stance*, in trying to understand its behavior [5]. For instance, it would be difficult to engage the robot in an intentional activity of *speech acts*, e.g., making a promise.

Now imagine a robot that has learned and is still learning human communicative behavior. Because the robot's intelligence has no blueprint and its repertoire is *incomplete* and *open* to extensions and modifications, taking a design stance is no longer necessary. To some degree, the robot would be able to understand and influence our mental states, like desires and beliefs; it would thus be able to predict and control our behavior, as well as to be predicted and controlled by us, to some degree. We would regard this robot as a *social being*, with whom we would cooperate and compete in our social activities.

The discussion above suggests that social intelligence should have an *ontogenetic history* that is open to further development and that the ontogeny should be similar to that of human interlocutors in a cultural and linguistic community [10]. Therefore, we are "bringing up" a robot in a physical and social environment equivalent to that experienced by a human infant. Section 2 introduces our infant robot, *Infanoid*, as an embodiment of a human infant with functionally similar innate constraints. Sections 3 to 5 describe how the robot acquires human communicative behavior through its interaction with human caregivers. The robot first acquires intentionality, then identifies with others mainly by means of joint attention, and finally understands the communicative intentions of others' behavior.

2. Infanoid, the Babybot

We begin with the premise that any socially communicative intelligence must have a *naturalistic embodiment*, i.e. a robot that is structurally and functionally similar to human sensori-motor systems. The robot interacts with its environment in the same way as humans do, implicitly sharing its experience with human interlocutors, and gets situated in the environment shared with humans [10].

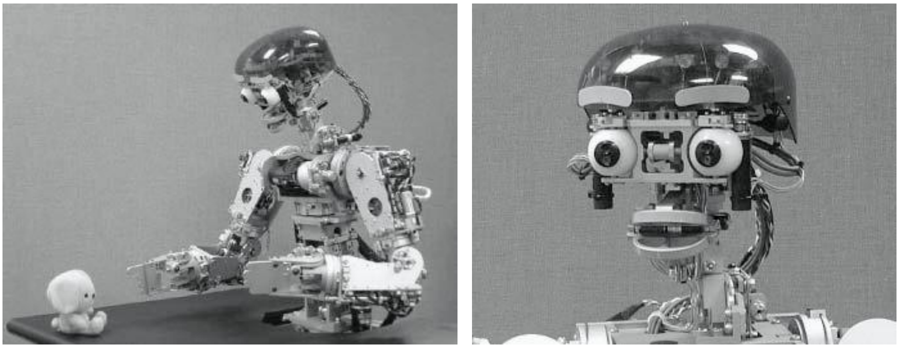


Figure 19.1. Infanoid, an upper torso humanoid (left), and its head (right).

Our robot, *Infanoid*, shown in Figure 19.1 (left), is being constructed as a possible naturalistic embodiment for communicative development. *Infanoid* possesses approximately the same kinematic structure of the upper body of a three-year-old human infant. Currently, 25 degrees of freedom (DOFs) — 7 in the head, 3 in the neck, 6 in each arm (excluding the hand), and 3 in the trunk

— are arranged in a 480-mm-tall upper body. *Infanoid* is mounted on a table for face-to-face interaction with a human caregiver sitting on a chair.

Infanoid has a foveated stereo vision head, as shown in Figure 19.1 (right). Each of the eyes has two color CCD cameras like those of *Cog* [3]; the lower one has a wide angle lens that spans the visual field (about 120 degrees horizontally), and the upper one has a telephoto lens that takes a close-up image on the fovea (about 20 degrees horizontally). Three motors drive the eyes, controlling their direction (pan and common tilt). The motors also help the eyes to perform a saccade of over 45 degrees within 100 msec, as well as smooth pursuit of visual targets. The images from the cameras are fed into massively parallel image processors (IMAP Vision) for facial and non-facial feature tracking, which enables real-time attentional interaction with the interlocutor and with a third object. In addition, the head has eyebrows with 2 DOFs and lips with 2 DOFs for natural facial expressions and lip-synching with vocalizations. Each DOF is controlled by interconnected MCUs; high-level sensori-motor information is processed by a cluster of Linux PCs.

Infanoid has been equipped with the following functions: (1) tracking a nonspecific human face in a cluttered background; (2) determining roughly the direction of the human face being tracked; (3) tracking objects with salient color and texture, e.g., toys; (4) pointing to or reaching out for an object or a face by using the arms and torso; (5) gazing alternately between the face and the object; and (6) vocalizing canonical babbling with lip-synching. Currently, we are working on modules for gaze tracking, imperfect verbal imitation, and so on, in order to provide *Infanoid* with the basic physical skills of 6-to-9-month-olds, as an initial stage for social and communicative development.

3. Being intentional

Communication is the act of sending and receiving physical signals from which the receiver derives the sender's *intention* to manifest something in the environment (or in the memory) so as to change the receiver's attention and/ or behavioral disposition [8]. This enables us to predict and control others' behavior to some degree for efficient cooperation and competition with others. It is easy to imagine that our species acquired this skill, probably prior to the emergence of symbolic language, as a result of the long history of the struggle for existence.

How do we derive intangible intentions from physically observable behavior of others? We do that by using *empathy*, i.e. the act of imagining oneself in the position of someone else, thereby understanding how he or she feels and acts, as illustrated in Figure 19.2. This empathetic process arouses in our mind, probably unconsciously, a mental state similar to that of the interlocutor. But, how can a robot do this? As well as being able to identify itself with the interlocutor, the robot has to be an *intentional being* capable of goal-directed spontaneous behavior by itself; otherwise, the empathetic process will not work.

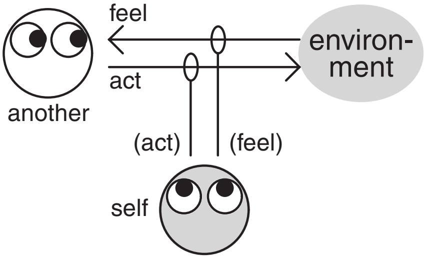


Figure 19.2. Empathy for another person's behavior.

In order to acquire intentionality, a robot should possess the following: (1) a *sensori-motor system*, with which the robot can utilize the affordance in the environment; (2) a *repertoire of behaviors*, whose initial contents are innate reflexes, e.g., grasping whatever the hand touches; (3) a *value system* that evaluates what the robot feels exteroceptively and proprioceptively; and (4) a *learning mechanism* that reinforces (positively or negatively) a behavior according to the value (e.g., pleasure and displeasure) of the result. Beginning with innate reflexes, which consist of a continuous spectrum of sensori-motor modalities, the robot explores the gamut of effective (profitable) *cause-effect* associations through its interaction with the environment. The robot is gradually able to use these associations spontaneously as *method-goal* associations. We have defined this as the acquisition of intentionality.

4. Being identical

To understand others' intentions, the intentional robot has to identify itself with others. This requires it to observe how others feel and act, as shown in Figure 19.2. *Joint attention* plays an important role in this understanding [1, 9], and *action capture* is also indispensable. Joint attention enables the robot to observe what others exteroceptively perceive from the environment, and action capture translates the observed action of others into its own motor program so that it can produce the same action or proprioception that is attached to that action.

4.1 Joint attention

Joint attention is the act of sharing each other's attentional focus.[1] It spotlights the objects and events being attended to by the participants of communication, thus creating a shared context in front of them. The shared context is a subset of the environment, the constituents of which are mutually manifested among the participants. The context plays a major role in reducing the computational cost of selecting and segmenting possible referents from the vast environment and in making their communicative interaction coherent.

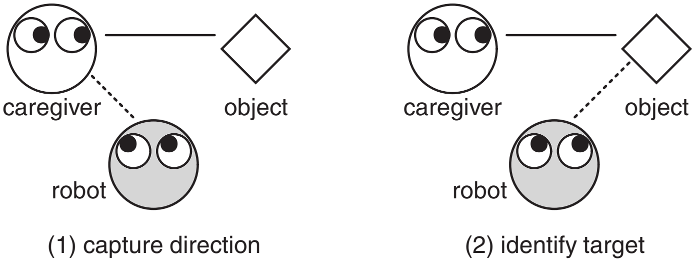


Figure 19.3. Creating joint attention with a caregiver.

Figure 19.3 illustrates how the robot creates and maintains joint attention with a caregiver. (1) The robot captures the direction of the caregiver's attention by reading the direction of the body, arms (reaching/pointing), face, and/ or gaze. (2) The robot does a search in that direction and identifies the object of the caregiver's attention. Occasionally the robot diverts its attention back to the caregiver to check if he or she is still attending to the object.

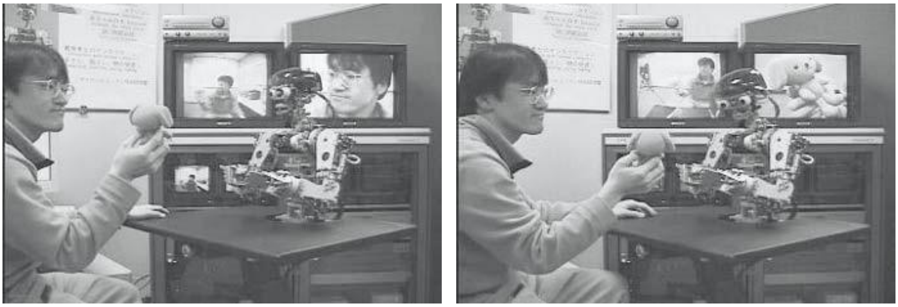


Figure 19.4. Infanoid engaging in joint attention.

As shown in Figure 19.4, *Infanoid* creates and maintains joint attention with the human caregiver. First, its peripheral-view cameras search for a human face in a cluttered video scene. Once a face is detected, the eyes saccade to the face and switch to the foveal-view cameras for a close-up image of the face. From this image, it roughly estimates the direction of the face from the spatial arrangement of the facial components. Then, *Infanoid* starts searching in that direction and identifies the object with salient color and texture like the toys that infants prefer.

4.2 Action capture

Action capture is defined as the act of mapping another person's bodily movements or postures onto one's own motor program or proprioception. This mapping connects different modalities; one *observes* another person's body exteroceptively (mainly visually) and *moves* or proprioceptively *feels* one's own body, as shown in Figure 19.5. Together with joint attention, action capture enables the robot to indirectly experience someone else's behavior, by translating the other person's behavior $\langle i, o\rangle$ into its own virtual behavior $\langle i', o'\rangle$, as illustrated in Figure 19.6.

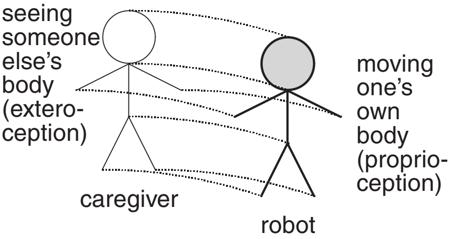

seeing someone else's body (exteroception)
moving one's own body (proprioception)
caregiver
robot

Figure 19.5. Mapping between self and another person.

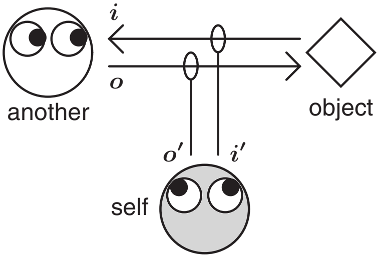


Figure 19.6. Indirect experience of another person's behavior.

A number of researchers have suggested that people are innately equipped with the ability to capture another person's actions; some of the mechanisms they have cited are *neonatal mimicry* [6] and *mirror neurons* [7]. Neonatal mimicry of some facial expressions is, however, so restricted that it does not fully account for our capability of whole-body imitation. Mirror neurons found in the pre-motor cortex of macaques activate when they observe someone doing a particular action and when they do the same action themselves. However, the claim that mirror neurons are the innate basis for action capture is not clear, since macaques do not imitate at all [4, 9].

To explain the origin of action capture, we assume that neonates possess *amodal* (or synesthetic) perception [2], in which both exteroception (of visual, tactile, etc.) and proprioception (of inner feelings produced from body postures and movements) appear in a single space spanned by dimensions such as spatial/temporal frequency, amplitude, and egocentric localization. This amodal perception would produce reflexive imitation, like that of facial expressions and head rotation. Beginning with quite a rough mapping, the reflexive imitation would get fine-tuned through social interaction (e.g., imitation play) with caregivers.

5. Being communicative

The ability to identify with others allows one to acquire empathetic understanding of others' intentions behind their behaviors. The robot ascribes the indirectly experienced behavior to the mental state estimated by using *self-reflection*. In terms of its own intentionality, self-reflection tells the robot the mental state that best describes the behavior. The robot then projects this mental state back onto the original behavior. This is how it understands others' intentions.

This empathetic understanding of others' intentions is not only the key to human communication, but also the key to *imitative learning*. Imitation is qualitatively different from emulation; while emulation is the reproduction of the same result by means of a pre-existing behavioral repertoire or one's own trial-and-error, imitation copies the intentional use of methods for obtaining goals [4, 9]. This ability to imitate is specific to *Homo sapiens* and has given the species the ability to share individual creations and to maintain them over generations, creating language and culture in the process [9].

Language acquisition by individuals also relies on the empathetic understanding of others' intentions. A symbol in language is not a label of referent, but a piece of high-potential information from which the receiver derives the sender's intention to manifest something in the environment [8]. The robot, therefore, has to learn the use of symbols to communicate intentions through identifying itself with others.

6. Conclusion

Our ontogenetic approach to social intelligence was originally motivated by the recent study of *autism* and related developmental disorders. Autism researchers have found that infants with autism have difficulty in joint attention and bodily imitation [1, 9], as well as in pragmatic communication. This implies that joint attention and action capture intertwine with each other, playing important roles in infants' development of social communication. Therefore, we have implemented in *Infanoid* the primordial capability of joint attention and are working on that of *action capture*.

Social intelligence has to have an ontogenetic history that is similar to that of humans and is open to further adaptation to the social environment; it also has to have a naturalistic embodiment in order to experience the environment in a way that is similar to humans'. Our ongoing attempt to foster *Infanoid* will tell us the prerequisites (nature) for and developmental process (nurture) of the artificial social beings that we can relate to.

Notes

1. Joint attention requires not only focusing on the same object, but also *mutual acknowledgement* of this sharing action. We assume that joint attention before "nine-month revolution" [9] is reflexive—therefore, without this mutual acknowledgement.

References

[1] S. Baron-Cohen. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA, 1995.

[2] S. Baron-Cohen. Is there a normal phase of synaesthesia in development? *Psyche*, 2(27), 1996. http://psyche.cs.monash.edu.au/v2/psyche-2-27-baron_cohen.html.

[3] R.A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. Williamson. The Cog project: building a humanoid robot. In C.L. Nehaniv, editor, *Computation for Metaphors, Analogy and Agents*, Lecture Notes in Computer Science, Vol. 1562, pages 52–87. Springer-Verlag, Berlin, 1998.

[4] R. Byrne. *The Thinking Ape: Evolutionary Origins of Intelligence*. Oxford University Press, 1995.

[5] D.C. Dennett. *The Intentional Stance*. MIT Press, Cambridge, MA, 1987.

[6] A. Meltzoff and M.K. Moore. Persons and representation: why infant imitation is important for theories of human development. In J. Nadel and G. Butterworth, editors, *Imitation in Infancy*, pages 9–35. Cambridge University Press, 1999.

[7] G. Rizzolatti and M.A. Arbib. Language within our grasp. *Trends in Neuroscience*, 21: 188–194, 1998.

[8] D. Sperber and D. Wilson. *Relevance: Communication and Cognition*. Harvard University Press, Cambridge, MA, 1986.

[9] M. Tomasello. *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge, MA, 1999.

[10] J. Zlatev. The epigenesis of meaning in human beings, and possibly in robots. *Minds and Machines*, 11: 155–195, 2001.

Chapter 20

PLAY, DREAMS AND IMITATION IN ROBOTA

Aude Billard

Computer Science Department, University of Southern California

Abstract Imitation, play and dreams are as many means for the child to develop her/his understanding of the world and of its social rules. What if we were to have a robot we could play with? What if we could through play and daily interactions, as we do with our children, be a model for it and teach it (what?) to be human-like? This chapter describes the Robota dolls, a family of small humanoid robots, which can interact with the user in many ways, imitating gestures, learning how to dance and learning how to speak.

1. Introduction

The title of this chapter is a wink to Swiss psychologist Jean Piaget and his book *Play, Dreams and Imitation in Childhood* [16]. For Piaget, imitation, play and dreams are as many means for the child to develop her/his understanding of the world and of its social rules. This chapter discusses the aspects of these behaviors which make them relevant to research on *socially intelligent agents* (SIA)[7].

Natural human-like interaction, such as imitation, speech and gestures are important means for developing likeable, *socially interactive robots*. This chapter describes the Robota dolls, a family of small humanoid robots. The Robota dolls can interact with the user in many ways, imitating gestures and learning from her/his teachings. The robots can be taught a simple language, little melodies and dance steps.

1.1 Play

Entertainment robotics (ER) is one of the many fields which will benefit from the development of socially intelligent agents. ER aims at creating playful autonomous creatures, which show believable animal-like behaviors [5]. Successful examples of such intelligent toys are, e.g., the *Tamagotchi*[1], the *Furbys*[2] and the *Sony Aibo* [12].

For psychologists (starting with Piaget), children's games are as much an educational tool as an entertainment device. Similarly, beyond the goal of making a successful toy, ER aims also at developing entertaining educational tools [8, 11]. An educational toy offers a challenge. It is such that, through play, the child explores new strategies and learns new means of using the toy. While this can be true of the simplest toy, such as a wooden stick (which can be used as a litt, a drill, a bridge), robotics faces the challenge to create a toy which is sophisticated while leaving sufficient freedom for the child imagination. This is made possible in two ways:

1) By making the robot's behavior (software) adaptable; the user takes part into the development of its creature (e.g. *Tamagotchi*, the video game *Creatures* [13], the baby dolls *My Real Baby*[3] and *My Dream Baby* [4]; the robot becomes more of a pet.

2) By offering flexibility in the design of the robot's body, e.g. *LEGO mindstorms*[5].

The Robota dolls have been created in this spirit. They have general learning abilities which allow the user to teach them a verbal and body (movement) language. Because they are dolls, the features of their humanoid body can be changed by the user (choice of skin color, gender, clothing).

1.2 Imitation

Following Piaget, a number of authors pointed out the frequent co-occurrence of imitation game during play, suggesting that "the context of play offers a special state of mind (relaxed and free from any immediate need) for imitative behavior to emerge" [15]. Imitation is a powerful means of social learning, which offers a wide variety of interaction. One can imitate gestures, postures, facial expressions, behaviors, where each of the above relates to a different social context. An interesting aspect of imitation in humans (perhaps as opposed to other animals) is that it is a bidirectional process [15]. Humans are capable to recognize that they are imitated. Imitation becomes also a means of teaching, where the demonstrator guides the imitator's reproduction.

Roboticists use imitative learning as a user-friendly means to teach a robot complex skills, such as learning the best path between two points [4, 6, 9], learning how to manipulate objects [14, 18], and, more generally, learning how to perform smooth, human-like movements by a humanoid robot [10, 17]. These efforts seek to enhance the robot's ability to interact with humans by providing it with natural, socially driven behaviors [7].

In the Robota dolls and other works [1, 2], we have exploited the robot's ability to imitate another agent, robot or human, to teach it a basic language. The imitation game between user and robot is a means to direct the robot's attention to specific perceptions of movement, inclination, orientation. The robot can then be taught words and sentences to describe those perceptions.

2. Robota

Figure 20.1 shows a picture of the two original Robota dolls. A commercial series of Robota dolls is now available[6] with different body features, including a purely robot-like (completely metallic) one.

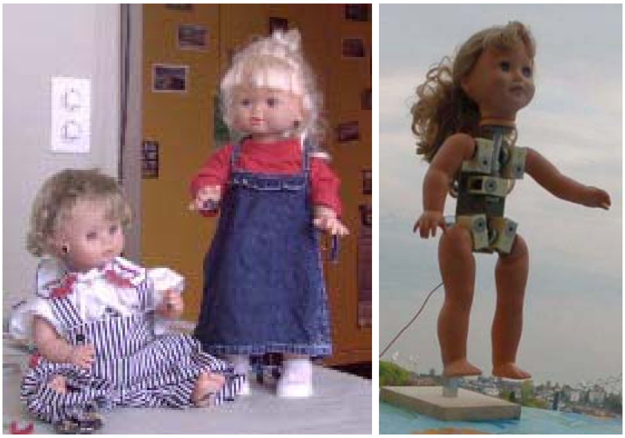


Figure 20.1. Left Picture: On the left, the first prototype of Robota doll made out of LEGO, and, on the right, the second prototype of Robota doll. Right Picture: The new commercial prototype (version Caucasian).

2.1 Technical specificities

These features are that of new series of Robota dolls.

General. The robot is 50 cm tall, weighting 500gr. The arms, legs and head of the robot are plastic components of a commercially available doll. The main body is a square box in transparent plexiglas, which contains the electronics and mechanics. It has an on-board battery of 30 minute duration.

Electronic. The behavior of the robot is controlled through a Kameleon K376SBC board[7], attached to the main body of the robot.

External interfaces. the robot connects to a keyboard (8 words), which can also be used as an electronic xylophone (8 notes), and a joystick (to control the movement). The robot can connect through a serial link to a PC (the code for the PC is written in C and C++ and runs both under linux and windows 95/98/2000. 96M RAM, Pentium II, 266MHz). A PC-robot interfacing program allows one to interact with the robot through speech and vision.

Motors. The robot is provided with 5 motors to drive separately the two arms, the two legs (forward motion) and the head (sideways turn). A prototype of motor system to drive the two eyes in coordinated sideways motion is under construction.

Imitation game with infra-red. The robot has 4 pairs of infra-red emitter/receptor to detect the user's hand and head movements. The sensors are mounted on the robot's body and the emitters are mounted on a pair of gloves and glasses which the user wear. The sensors on the robot's body detect the movement of the emitters on the head and hands of the user. In response to the user's movement, the robot moves (in mirror fashion) its head and its arms, as shown in Figure 20.2 (left).

Imitation game with camera. A wireless CCD camera (30MHZ) attached to a PC tracks optical flow to detect vertical motion of the left and right arms of the instructor. The PC sends via the serial link the position of each of the instructor's arm to direct the mirror movement in the robot (Figure 20.2, right).

Other Sensors. The robot is provided with touch sensors (electrical switches), placed under the feet, inside the hands, on top of the head and in the mouth, a tilt sensor which measures the vertical inclination of the body and a pyroelectric sensor, sensitive to the heat of human body.

Speech. Production and recognition of speech is provided by ELAN synthesizer[8] and speech processing software from Viavoice (in French) and Dragon (in English). Speech is translated into ordered strings of words (written language).

2.2 Software: Behavioral capabilities

"Baby behaviors". The Robota doll can engage in a simple interaction with the user by demonstrating baby-like behaviors, which requires the user to "take care" of the robot. These are built-in behaviors, implemented as a set of internal variables (happiness, tiredness, playfulness and hungriness) which vary over time. For a given set of values, the robot will start to cry, laugh, sing or dance. In a sad mood, it will also extend the arms for being rocked and

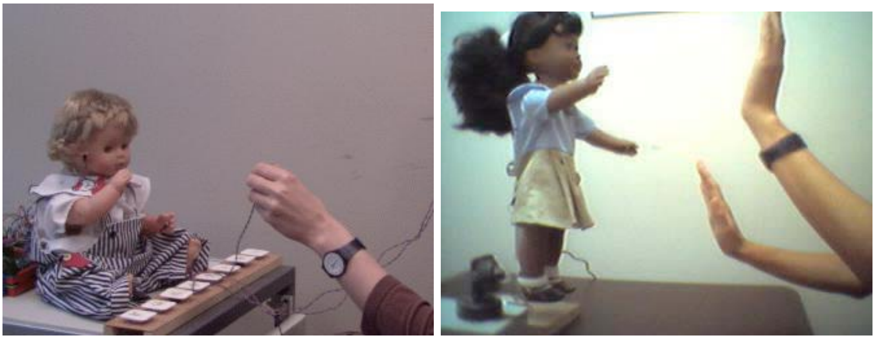


Figure 20.2. Left: The teacher guides the motions of Robota using a pair of glasses holding a pair of IR emitter. The glasses radiation which can be picked up by the robot's "earrings" IR receptors. Right: Robotina, the latino version of Robota mirrors the movements of an instructor by tracking the optical flow created by the two arms moving in front of the camera located on the left side of the robot.

babble to attract attention. In response to the care-giver's behavior the "mood" of the robot varies, becoming less hungry when fed, less tired when rocked and less sad when gently touched.

Learning behavior. The robot is endowed with learning capacities provided by an artificial neural network [4], which has general properties for learning complex time series. The algorithm runs both on the PC interface and on-board of the robot. When using the PC speech interface, the user can teach the robot a simple language. The robot is taught by using complete sentences ("You move your leg", "I touch your arm", "You are a robot"). After several teachings, the robot learns the meaning of each word by extracting the invariant use of the same string in the sentences. It can learn verbs ('move', 'touch'), adjectives ('left', 'right') and nouns ('foot', 'head'). In addition, the robot learns some basic syntactic rules by extracting the precedence of words in the sentence (e.g. the verb "move" comes always before the associated noun "legs"). Once the language is learned, the robot responds to the user, by speaking new combinations of words for describing its motions and perceptions.

The learning algorithm running on-board of the robot allows learning of melodies and of simple word combinations (using the keyboard) and learning of dance movement (using the imitation game) by association of movements with melodies.

3. Dreams

To conclude this chapter, I wish to share with you my dreams for Robota and my joy in seeing some of those being now realized.

3.1 A toy and educational tool

An important motivation behind the creation of the first Robota doll was to make it an appealing show-case of Artificial Intelligence techniques. This wish is now realized thanks to the museum *La cité des sciences et de l'industrie*[9], which will be presenting it from November 2001 to March 2003.

I also wished to create a cute, but interesting toy robot. In order to achieve this, I provided the robot with multimedia type of interactions. In spring 1998, tests with children of 5 and 6 years old showed the potential of the system as a game for children [3]. The children showed pleasure when the robot reacted to their movements. The robot would respond to the children touching specific parts of its body, by making small movements or little noises. It would mimic the child's head and arm movements. Because imitation is a game that young children like to play with each other and their parents, it was easy for them to understand that they could interact with the robot in this way. The children managed to teach the robot some words part of their every-day vocabulary (e.g. *food, hello, no*) and showed satisfaction when the robot would speak the words back.

Another important wish was that the robot would be useful. In this spirit, I have sought collaboration with educators and clinicians. One key feature of the robot as an educational tool is that the level of complexity of the game with Robota can be varied. One can restrict oneself to only interact with the built-in behaviors of the robot (a baby-like robot). The learning game can be restricted to learning only music patterns (using the musical keyboard), dance patterns, or speech.

This lead to the idea of using the game with Robota (by exploiting the different degrees of complexity) to train and possibly test (in the case of retarded children and, e.g., for evaluating the deepness of autism) the child's motor and linguistic competences. In October 1999, as part of Kerstin Dautenhahn's Aurora project[10], the first prototype of Robota was tested at Radlett Lodge School with a group of children with autism. Although the interactions were not formally documented, observations showed that the children showed great interest in the robot. Consistent with general assumptions about autism, they showed interest in details of the robot (e.g. eyes, cables that were visible etc.). In collaboration with Kerstin Dautenhahn, further tests will be carried out to evaluate the possible use of the robot in her projects.

Current collaboration with Sharon Demuth, clinician, and Yvette Pena, director of the USC premature infant clinic (Los Angeles) conducts pilot studies to evaluate the use of the robot with premature children. The idea there is that robot would serve as an incentive for the child to perform its daily necessary exercises, in order to overcome its motor weaknesses, as well as its verbal delay.

My dream is now that these studies will lead to some benefits for the children involved, if only to make them smile during the game.

Acknowledgments

Many thanks to Jean-Daniel Nicoud, Auke Ijspeert, Andre Guignard, Olivier Carmona, Yuri Lopez de Meneses and Rene Beuchat at the Swiss Institute of Technology in Lausanne (EPFL) and Alexander Colquhun and David Wise at the University of Edinburgh for their support during the development of the electronic and mechanical parts of the first prototypes of the Robota dolls. Many thanks to Marie-Pierre Lahalle at the CSI Museum, Kerstin Dautenhahn, Sharon Demuth and Yvette Pena for their support in the diverse projects mentioned in this paper.

Notes

1. www.bandai.com.
2. www.furby.com.
3. www.irobot.com.
4. www.mgae.com.
5. mindstorms.lego.com.
6. www.Didel.com, SA, CH.
7. www.k-team.com.
8. www.elan.fr.
9. CSI, Paris, www.csi.fr.
10. www.aurora-project.com.

References

[1] A. Billard. Imitation: a means to enhance learning of a synthetic proto-language in an autonomous robot. In C. Nehaniv and K. Dautenhahn, editors, *Imitation in Animals and Artifacs*. MIT Press, Cambridge, MA, 2002 (In Press).

[2] A. Billard and K. Dautenhahn. Experiments in social robotics: grounding and use of communication in autonomous agents. *Adaptive Behavior, special issue on simulation of social agents*, 7(3/4): 415–438, 1999.

[3] A. Billard, K. Dautenhahn, and G. Hayes. Experiments on human-robot communication with robota, an imitative learning and communicating doll robot. In K. Dautenhahn and B. Edmonds, editors, *Proceedings of Socially Situated Intelligence Workshop* held within the *Fifth Conference on Simulation of Adaptive Behavior (SAB'98)*. Centre for Policy Modelling technical report series: No. CPM–98–38, Zurich, Switzerland, 1998.

[4] A. Billard and G. Hayes. Drama, a connectionist architecture for control and learning in autonomous robots. *Adaptive Behavior*, 7(1): 35–64, 1999.

[5] J. Cassell and H. Vilhjálmsson. Fully embodied conversational avatars: Making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems*, 2(1):45–64, 1999.

[6] K. Dautenhahn. Getting to know each other – artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 16:333–356, 1995.

[7] K. Dautenhahn. Embodiment and interaction in socially intelligent life-like agents. In C.L. Nehaniv, editor, *Computation for Metaphors, Analogy and Agent*, Lecture Notes in Artificial Intelligence, Volume 1562, pages 102–142. Springer, Berlin and Heidelberg, 1999.

[8] K. Dautenhahn. Robots as social actors: Aurora and the case of autism. In *Proc. CT99, The Third International Cognitive Technology Conference*. August, San Francisco, CA, 1999.

[9] J. Demiris and G. Hayes. Imitative learning mechanisms in robots and humans. In *Proceedings of the 5th European Workshop on Learning Robots*, pages 9–16. Bari, Italy, July 1996. Also published as Research Paper No 814, Dept. of Artificial Intelligence, University of Edinburgh, UK, 1996.

[10] Y. Demiris and G. Hayes. Imitation as a dual-route process featuring predictive and learning components: A biologically-plausible computational model. In C. Nehaniv and K. Dautenhahn, editors, *Imitation in Animals and Artifacs*. MIT Press, Cambridge, MA, 2002 (In Press).

[11] A. Druin, B. Bederson, A. Boltman, A. Miura, D. Knotts-Callahan, and M. Platt. Children as our technology design partners. In A. Druin, editor, *The Design of Children's Technology*. The Morgan Kaufmann Series in Interactive Technologies, 1998.

[12] M. Fujita and H. Kitano. Development of an autonomous quadruped robot for robot entertainment. *Autonomous Robots*, 5(1): 7–18, 1998.

[13] S. Grand. Creatures: an exercise in creation. *IEEE Intelligent Systems*, 12(4): 19–24, 1997.

[14] M.I. Kuniyoshi and I. Inoue. Learning by watching: Extracting reusable task knowledge from visual observation of human performance. *IEEE Transactions on Robotics and Automation*, 10(6): 799–822, 1994.

[15] Á. Miklósi. The ethological analysis of imitation. *Biological Review*, 74:347–374, 1999.

[16] J. Piaget. *Play, Dreams and Imitation in Childhood.* Norton, New York, 1962.

[17] S. Schaal. Learning from demonstration. *Advances in Neural Information Processing Systems*, 9:1040–1046, 1997.

[18] S. Schaal. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 3(6):233–242, 1999.

Chapter 21

EXPERIENCES WITH SPARKY, A SOCIAL ROBOT

Mark Scheeff, John Pinto, Kris Rahardja, Scott Snibbe and Robert Tow
All formerly of Interval Research Corporation[*]

Abstract In an effort to explore human response to a socially competent embodied agent, we have a built a life-like teleoperated robot. Our robot uses motion, gesture and sound to be social with people in its immediate vicinity. We explored human-robot interaction in both private and public settings. Our users enjoyed interacting with Sparky and treated it as a living thing. Children showed more engagement than adults, though both groups touched, mimicked and spoke to the robot and often wondered openly about its intentions and capabilities. Evidence from our experiences with a teleoperated robot showed a need for next-generation autonomous social robots to develop more sophisticated sensory modalities that are better able to pay attention to people.

1. Introduction

Much work has been done on trying to construct intelligent robots but little of that work has focused on how human beings respond to these creatures. This is partly because traditional artificial intelligence, when applied to robotics, has often focused on tasks that would be dangerous for humans (mine clearing, nuclear power, etc.). Even in the case of tasks in which humans are present, people are mostly seen as obstacles to be avoided. But what if we conceive of a class of robots that are explicitly social with humans, that treat humans not as obstacles, but as their focus? There are at least two sides to this problem that need studying: first, how do you construct a socially competent robot and, second, how do people respond to it. Our work has focused on studying the latter question, human response to a socially competent robot.

To that end, we have constructed a robot, Sparky, whose purpose is to be social with humans in its vicinity. Since we are studying human response, we have *not* tried to solve the problem of generating reasonable autonomous action. Rather, we have built a teleoperated device, and manifested a degree of social intelligence which we believe could be accomplished autonomously in the near, though not present, future.

Our studies were a broad ranging exploration that asked open-ended questions. Would people find Sparky compelling or disturbing? What behaviors would people exhibit around the robot? What new skills does a robot need to develop when it is in a social setting (and what skills can it forget)? We hope that our findings can help to guide the development of future robots that either must or would like to be social with humans. We also hope that our work points to the potential for interface devices that use a physical system (a body) as a way to communicate with users.

2. Prior Work

In searching for inspiration in creating life-like characters, we first looked towards the principles of traditional animation and cartooning [13, 5]. The computer graphics community has also explored many ways of creating realistic, screen-based, animated characters [1, 11]. We ended up using Ken Perlin's Improv system [7] as the foundation for our approach to movement.

Masahiro Mori has written eloquently on the perils of building a robot that resembles a living creature *too much*. His point, that cartoons or simplified representations of characters are generally more acceptable to people than complicated "realistic" representations, became an important tool in making our design decisions (adapted from [9]).

The emerging field of affective computing also provided motivation and justification for our work [8]. In an example of this type of endeavor, Breazeal [3, 2] has built an animated head, called Kismet, that can sense human affect through vision and sound and express itself with emotional posturing. Darwin's timeless work [4] inspired us to use a face on our robot.

Lastly, Isbister [6] has written an excellent discussion on the difference between traditional notions of intelligence, which emphasize the construction of an accurate "brain", and the idea of perceived intelligence, which emphasizes the perceptions of those who experience these artificial brains. This work helped us to understand how users saw intelligence in unfamiliar people or devices.

3. Our Robot, Sparky

Sparky is about 60cm long, 50cm high and 35cm wide (Figure 21.1). It has an expressive face, a movable head on a long neck, a set of moving plates on its back and wheels for translating around the room. A remote operator manifests the personality we have constructed for Sparky in a manner similar to giving directions to an actor on a stage: some movements are set explicitly and then a global emotional state is set. Sparky's onboard computer interprets these commands to drive all 10 degrees of freedom. Sparky appears autonomous to those around it.

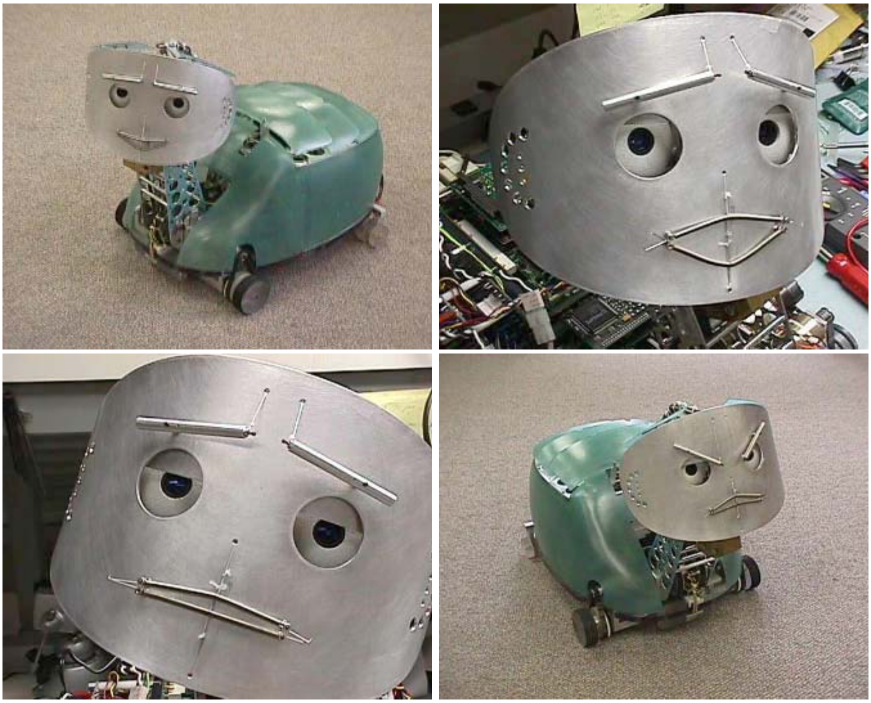


Figure 21.1. Sparky showing several emotions and postures.

During operation, Sparky is usually a friendly robot, approaching anyone in the vicinity while smiling and making an occasional happy utterance. Sometimes, though, our operator will command Sparky to act sad, nervous or fearful. If our robot suffers abuse, the operator can switch it into the "angry" emotion and, in extreme circumstances, even charge the abuser head on. Sparky can express nine different emotional states: neutral, happy, sad, angry, surprised, fearful, inquisitive, nervous, and sleepy.

Because of the way we control our robot, Sparky makes extensive use of its body. It will often track humans' eyes, crane its next backwards and forwards and mimic people's motions. It can even raise the hackles on its back, a gesture reminiscent of a cat.

Sparky is always moving and shifting its joints, much like a living creature. The type and amount of ambient motion is a result of the emotional state set by the operator and is generated automatically. We have written special software [12] based on Perlin's Improv system [7] to do this.

We can also cue Sparky to make vocalizations, which sound something like muffled speech combined with a French horn. Just as in the case of ambient motion, the affective content of each sound is correlated to Sparky's emotional state. There are several sounds available in each state.

A more comprehensive description of the robot is provided in our previous work [10].

4. Observing Sparky and People

To explore our research questions two venues were chosen in which to explore human-robot interaction, one in the lab and the second in public.

In the Lab. Thirty external subjects were recruited for 17 trials in our internal lab (singles and dyads). Approximately 50% of subjects were between ages 8–14, 13% were 19–30, 17% were 35–45 and 20% were over age 65. There was an even mix of genders. Subjects answered several background questions, interacted with the robot for about 15 minutes, and then discussed the experience with the interviewer in the room. Interactions between the robot and the subject were necessarily chaotic; we tried simply to react reasonably to the subject's actions while still manifesting the personality we have described above.

In Public. Tests were conducted 2–3 hours a day for six days at an interactive science museum. The robot was released for an hour at a time to "wander" in an open area. There were no signs or explanations posted.

5. Reactions

Reactions are grouped into three categories. In "Observed behavior" we report on what users did with the robot. In "Interview response" we cover the feedback they gave to the interviewer in lab testing. Finally, in "Operating the robot" we report on what the operators experienced.

5.1 Observed behavior

Children were usually rapt with attention and treated the robot as if it were alive. Young children (4–7ish) tended to be very energetic around the robot (giddy, silly, etc.) and had responses that were usually similar regardless of gender. They were generally very kind to Sparky. Occasionally, a group of children might tease or provoke Sparky and we would then switch into a sad, nervous, or afraid state. This provoked an immediate empathetic response.

Older children (7ish to early teens) were also engaged but had different interaction patterns depending on gender. Older boys were usually aggressive towards Sparky. Boys often made ugly faces at the robot and did such things as covering the eyes, trapping it, pushing it backwards and engaging in verbal abuse. Switching the robot to a sad, nervous or fearful emotional state actually increased the abuse. Moving to an angry and aggressive emotional state seemed to create a newfound respect.

Older girls were generally gentle with the robot. Girls often touched the robot, said soothing things to it, and were, on occasion, protective of the robot. If an older girl did provoke Sparky a little and it switched into a sad emotion, empathy was the result. It should be noted that although the responses for older boys and girls were stereotypical, exceptions were rare.

Most adult interaction was collected in our lab. Adults tended to treat the robot like an animal or a small child and generally gave the impression that they were dealing with a living creature. Compared to children, they were less engaged. Gender wasn't a significant factor in determining adult responses. Response to Sparky's emotional palette was similar to the results with young children and older girls.

In the lab, most adults quickly began to play with the robot. Some however, were clearly unsure what to do. Many of these people eventually began to experiment with the robot (see below).

As we reviewed our data, we found that certain behaviors showed up quite often. These are catalogued below.

- Many subjects touched the robot. This behavior was more prevalent in young people, but was still common in adults as well. Once again, older children had responses that varied with gender. Boys were rougher, more likely to push it or cover its face. Girls tended to stroke and pet the robot. Adult touching was more muted and not dependent on gender.

- Subjects talked to the robot quite a bit. They sometimes interpreted the robot for other people and "answered" the robot when it made vocalizations. They often heard the robot saying things that it hadn't and assumed that its speech was just poor, rather than by design. Users often asked several questions of the robot, even if the robot ignored them. The most common question was "what's your name?"

- It was very common for subjects to mimic some portion of the robot's motion. For instance, if the robot moved its head up and down in a yes motion, subjects often copied the gesture in time with it. They also copied the extension and withdrawal of the head and its motion patterns.

- When a subject first engaged with the robot, s/he usually did so in one of two ways. The active subject stood in front of the robot and did something that might attract attention (made a face, waved, said something). The passive subject stood still until the robot acknowledged the subject's presence. Essentially, the passive subject waited to be acknowledged by the robot, while the active subject courted a response.

- Some subjects, mostly adults, spent time trying to understand the robot's capabilities better. For instance, subjects would snap their fingers to see if the robot would orient to the sound, or they would move their hands and bodies to see if the robot could follow them.

5.2 Interview response

Formal subject feedback was collected in the lab testing. Overall, subjects liked interacting with the robot and used such adjectives as "fun", "neat", "cool", "interesting" and "wild". The responsiveness of the robot in its movement and emotions was cited as compelling. In particular, subjects often mentioned that they liked how the robot would track them around the room and even look into their eyes. Subjects commented that the robot reminded them of a pet or a young child.

For some, primarily adults, motivation was a confusing issue. Though they typically could understand what the robot was expressing, subjects sometimes did not know *why* the robot acted a certain way. Also, vocalizations of the robot were not generally liked, though there were exceptions. Most found Sparky's muffled tone frustrating as they expected to be able to understand the words, but couldn't (by design, ironically).

5.3 Operating the robot

One of our project goals was to understand what new skills a social robot would need to learn. We therefore noted what our operators did as well.

Though it was not surprising, operators consistently got the best engagement by orienting the robot to the person. The robot's face pointed to the human's face and, moreover, we consistently found it valuable to look directly into the human's eyes. Being able to read the basic affect of human faces was also valuable.

Operators also found themselves having to deal with the robot's close proximity to many quickly moving humans. Users expected Sparky to know that they were there. For instance, if they touched Sparky somewhere, they expected it to know that and act accordingly (not move in that direction, turn its head to look at them, etc.).

6. Discussion and Conclusions

Users enjoyed interacting with Sparky and treated it as a living thing, usually a pet or young child. Kids were more engaged than adults and had responses that varied with gender and age. No one seemed to find the robot disturbing or inappropriate. A friendly robot usually prompted subjects to touch the robot, mimic its motions and speak out loud to it. With the exception of older boys, a sad, nervous or afraid robot generally provoked a compassionate response.

Our interactions with users showed a potential need for future (autonomous) social robots to have a somewhat different sensory suite than current devices. For instance, we found it very helpful in creating a rich interaction to "sense" the location of bodies, faces and even individual eyes on users. We also found it helpful to read basic facial expressions, such as smiles and frowns. This argues for a more sophisticated vision system, one focused on dealing with people. Additionally, it seemed essential to know where the robot was being touched. This may mean the development of a better artificial skin for robots. If possessed by an autonomous robot, the types of sensing listed above would support many of the behaviors that users found so compelling when interacting with a teleoperated Sparky.

Fortunately, there are some traditional robotic skills that Sparky, if it were autonomous, might not need. For instance, there was no particular need for advanced mapping or navigation and no need, at least as a purely social creature, for detailed planning. A robot that could pay attention to people in its field of view and had enough navigation to avoid bumping into objects would probably do quite well in this human sphere. Even if future robots did occasionally bump into things or get lost, it shouldn't be a problem: Sparky was often perceived as acting reasonably even when a serious control malfunction left it behaving erratically. When the goal is to be perceived as "intelligent", there are usually many acceptable actions for a given situation. Though it will be challenging to build these new social capabilities into mobile robots, humans are perhaps a more forgiving environment than roboticists are accustomed to.

We close on a speculative, and perhaps whimsical, note. Users interacted with Sparky using their bodies and, in turn, received feedback using this same, nearly universal, body language. This left us thinking not only of robots, but also of the general question of communication in computer interfaces. What if these human-robot interactions were abstracted and moved into other realms and into other devices? For instance, the gestures of head motion and gaze direction could map readily to a device's success at paying attention to a user. Similarly, Sparky could intuitively demonstrate a certain energy level using its posture and pace. Could another device use this technique to show its battery state? Though our research didn't focus on these questions, we believe this could be fertile ground for future work.

Notes

[*] Contact author: mark@markscheeff.com.

References

[1] B. Blumberg and T. Galyean. Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. *Computer Graphics*, 30(3): 47–54, 1995.

[2] C. Breazeal. Designing Sociable Machines: Lessons Learned. *This volume*.

[3] C. Breazeal and B. Scassellati. Infant-like Social Interactions Between a Robot and a Human Caretaker. *Adaptive Behavior*, 8(1): 49–74, 2000.

[4] C. Darwin. *The Expression of the Emotions in Man and Animals*. Oxford University Press, Oxford, UK, 1872.

[5] J. Hamm. *Cartooning the Head and Figure*. Perigee Books, New York, 1982.

[6] K. Isbister. *Perceived Intelligence and the Design of Computer Characters*. M.A. thesis, Dept. of Communication, Stanford University, Stanford, CA, 1995.

[7] K. Perlin and A. Goldberg. Improv: A System for Scripting Interactive Actors in Virtual Worlds. In *Proceedings of Siggraph 1996*, pages 205–216. ACM Press, New York, 1996.

[8] R. Picard. *Affective Computing*. The MIT Press, Cambridge, MA, 1997.

[9] J. Reichard. *Robots: Fact, Fiction and Prediction*. Penguin Books, London, 1978.

[10] M. Scheeff, J. Pinto, K. Rahardja, S. Snibbe, and R. Tow. Experiences with Sparky, A Social Robot. In *Proceedings of the 2000 Workshop on Interactive Robotics and Entertainment*, pages 143–150. Carnegie Mellon University, Pittsburgh, Pennsylvania, April 30 – May 1, 2000.

[11] K. Sims. Evolving Virtual Creatures. In *Proceedings of SIGGRAPH 1994*, pages 15–22. ACM Press, New York, 1994.

[12] S. Snibbe, M. Scheeff, and K. Rahardja. A Layered Architecture for Lifelike Robotic Motion. In *Proceedings of the 9th International Conference on Advanced Robotics*. Japan Robotics Association, Tokyo, October 25–27, 1999.

[13] F. Thomas and O. Johnston. *The Illusion of Life: Disney Animation*. Hyperion, New York, 1981.

Chapter 22

SOCIALLY SITUATED PLANNING

Jonathan Gratch
USC Institute for Creative Technologies

Abstract This chapter describes techniques to incorporate richer models of social behavior into deliberative planning agents, providing them the capability to obey organizational constraints and engage in self-interested and collaborative behavior in the context of virtual training environments.

1. Socially Situated Planning

Virtual environments such as training simulators and video games do an impressive job at modelling the physical dynamics but fall short when modelling the social dynamics of anything but the most impoverished human encounters. Yet the social dimension is at least as important as graphics for creating an engaging game or effective training tool. Flight simulators can accurately model the technical aspects of flight but many aviation disasters arise from social breakdowns: poor crew management, or the effects of stress and emotion on decision-making. Perhaps the biggest consumer of simulation technology, the U.S. military, identifies unrealistic human and organizational behavior as a major limitation of existing simulation technology [5].

There are many approaches to modelling social behavior. Socially-situated planning focuses on the problem of generating and executing plans in the context of social constraints. It draws inspiration from the shared-plans work of Grosz and Kraus [3], relaxes the assumption that agents are cooperative and builds on more conventional artificial intelligence planning techniques. Social reasoning is modelled as an additional layer of reasoning atop a general purpose planning. The planner handles task-level behaviors whereas the social layer manages communication and biases plan generation and execution in accordance with the social context (as assessed within this social layer). In this sense, social reasoning is formalized as a form of meta-reasoning.

Social Assessment: To support a variety of social interactions, the social reasoning layer must provide a model of the social context. The social situation is described in terms of a number of static and dynamic features from a particular agent's perspective. Static features include innate properties of the character being modelled (social role and a small set of "personality" variables). Dynamic features are derived from a set of domain-independent inference procedures that operate on the current mental state of the agent. These include the set of current communicative obligations, a variety of relations between the plans in memory (your plans threaten my plans), and a model of the emotional state of the agent (important for its communicative role).

Planning: One novel aspect of this work is how the social layer alters the planning process. Grosz and Kraus show how meta-level constructs like commitments can act as constraints that limit the planning process in support of collaboration (for example, by preventing a planner from unilaterally altering an agreed upon joint plan). We extend this to model a variety of "social stances" one can take towards other individuals beyond purely collaborative relationships. Thus, the social layer can bias planning to be more or less considerate to the goals of other participants and model power relationships between individuals.

Communication: Another key aspect of social reasoning is the ability to communicate socially appropriate information to other agents in the virtual environment. As with many approaches to social reasoning, the social layer provides a set of speech acts that an agent can use to convey or request information. Just as plan generation should differ depending on the social situation, the use of speech acts must be similarly biased. A commanding officer in a military operation would communicate differently and under different contexts than her subordinates.

Social Control Programs: Rather than attempting to formalize some specific rules of social behavior, we've adopted the approach of providing what is essentially a programming language for encoding the reasoning of the social layer. This language provides a set of inference procedures and data structures for representing an agent's social state, and it provides a set of control primitives that initiate communicative acts and alter the behavior of the task-level planning system. A simulation developer has a great deal of latitude in how they write "social control programs" that inform an agent's social-level reasoning. The strong constraint imposed by this language is that social reasoning is forced to operate at a meta-level. The control primitives treat plans as an indivisible unit. An agent can have multiple plans "in mind" and these can be communicated and treated differently by the planner, but the social-layer cannot manipulate or refer to the contents of these plans directly. This concept will be made clearer in the discussion below. These social control programs can be viewed as defining a finite state machine that changes the state of the set of control primitives based on features of the social context. In the examples in this chapter this state machine is defined in terms of a set of condition action rules, although in one application these state transitions have been formalized in terms of STRIPS-style planning operators and the social-program actually synthesized by the planning system [2].

2. Illustration

This approach has been used to model the behavior of military organizations [2] but the following contrived example provides a clearer view of the capabilities of the system. In this example, two synthetic characters, Jack and Steve, interact in the service of their own conflicting goals. The interaction is determined dynamically as the agents interact with each other, but is also informed by static information (e.g. the social stance they take towards one another).

These agents are embodied in a distributed virtual environment developed by Rickel and Johnson [6] that provides a set of perceptual, communicative and motor processes to control 3D avatars (see figure 22.1) that gesture and exhibit facial expressions. The agents share task knowledge encoded as STRIPS-style operators. They know how to drive vehicles to different locations, how to surf, and how to buy lottery tickets. They also have individual differences. They have differing goals, have varying social status and view their relationship with each other differently.

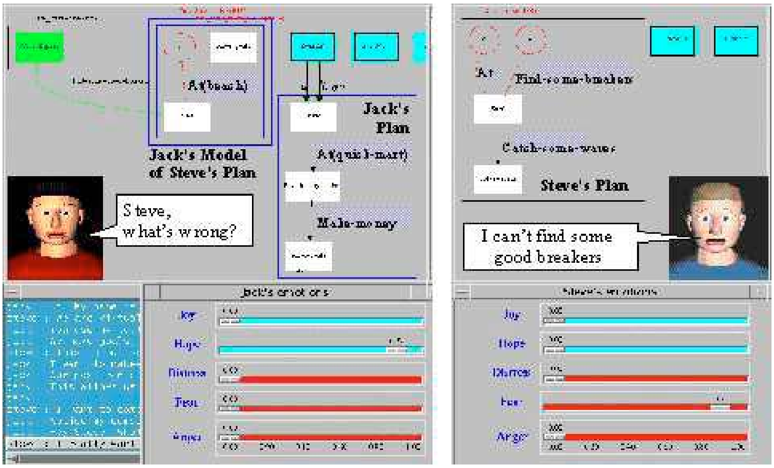


Figure 22.1. The 3D avatars Jack and Steve.

Jack's goal is to make money. Steve wants to surf. Both agents develop different plans but have to contend with a shared resource (a car). Besides performing task level actions, the agents engage in speech acts and generate gestures, facial expressions, and affective speech modulation based on properties of their social state.

The following are annotated traces of the system where the only difference is a change in the "personality" of the one agent. In the first trace, Steve treats Jack rudely, in the second, he treats him fairly. Text generation is extremely simplistic: agents actually communicate through a more stylised plan-communication language.

Rude Interaction:

Jack: I want to make-some-big-money. *[Creates a new plan containing this goal – enter car, drive to market, exit car, buy lotto ticket. Looks concerned, scratches his head, then, after devising a plan looks hopeful.]*

Steve: I want to catch-some-waves. *[Creates a new plan containing this goal. Looks concerned, scratches head. Surfing is important to Steve but a plan cannot be generated due to an unestablished "wave-location" precondition.]*

Jack: [*"sees" Steve's distress and requests information.]* Hey Steve, what's wrong?

Steve: *[Locates the feature in memory generating the most intense negative emotional excitation (see [1]). Communicates the associated plan in a distressed tone of voice.]* I want to catch some waves but can't find any good breakers.

Jack: *[Incorporates Steve's plan into memory and locates relevant information. Jack was previously told of a "plan" that establishes Steve's blocked subgoal]* Steve, does it help that someone did say there's some great waves near the pier?

Steve: *[Incorporates Jack's information. Derives a plan to go surfing. Looks hopeful.]*

Jack: *[Perceives Steve's change in expression and seeks to confirm his expectation that the information he provided helped Steve.]* So that information helped?

Steve: *[Handles Jack's information request.]* Yes Jack. I plan to drive the car to the beach, then I plan to surf-my-brains-out.

Jack: *[Incorporates Steve's revised plan and finds a conflict with his own plans. Based on personality, Jack attempts to negotiate a fair solution.]* Wait a second. Our plans conflict. I plan to drive the car to the quicky-mart then I plan to buy a-lottery-ticket.

Steve: *[Incorporates Jack's plan and recognizes the same interaction. Based on personality model, Steve responds to interaction differently. He devises a plan that satisfies his own goals without regard to any conflicts it may introduce in Jack's plans. Steve exits stage right.]* Later dude, I'm driving the car to the beach.

Jack: *[Perceives that car has departed without him. Looks angry. Says in angry voice:]* I want to kill-my-roommate.

Cooperative Interaction: Jack: *[Incorporates Steve's revised plan and finds a conflict with his own plans. Based on personality, Jack attempts to negotiate a fair solution.]* Wait a second. Our plans conflict. I plan to drive the car to the-quicky-mart then I plan to buy a-lottery-ticket.

Steve: *[Incorporates Jack's plan and recognizes the same interaction. Based on Steve having somewhat lower social status, he takes the initiative in repairing the conflict.]* Well, I could

change my plans. *[Looks concerned, scratches head, then devises a possible joint plan.]* I have a suggestion. Could you drive the car to the-quicky-mart with-me then I could drive the car to the beach. *[Note that neither agent has been given the goal of returning home.]*
Jack: *[Adds Steve's suggested joint plan, determines that it is consistent with his own, and agrees to form a joint commitment to the shared plan.]* Sounds good to me.

3. Social Control Programs

A small change in an agent's static social state can result in a dramatic change in behavior because reasoning at the social level is highly leveraged. Social reasoning is conditioned on dynamic social features that encapsulate a good deal of domain-independent inference and social control primitives allow for considerable differences in how plans are generated and executed at the base level. Social reasoning is represented as a set of condition actions rules that operate at this meta-layer. Social state components serve as the conditions for these social rules whereas control primitives define the space of possible actions.

3.1 Social State

An agent's social state is composed of dynamic and static components. Dynamic components are further divided into communicative state, plan state, and emotional state.

Communicative State: The communicative state tracks what information has been communicated to different agents and maintains any communicative obligations that arise from speech acts. When Steve communicates a plan to Jack, Steve's social layer records that Jack knows this plan, and persists in knowing it until Steve's planning layer modifies it, at which point Steve's social layer records that Jack's knowledge is out of date. If Jack requests Steve's current plans, the social layer creates communicative obligations: the fact that Steve owes Jack a response is recorded in each agent's social layer (though whether Steve satisfies this obligation is up to Steve's social control program).

Plan State: At the base-level planning layer, all activities that an agent is aware of (whether they come from its own planning or are communicated from outside) are stored in a single plan network, allowing the planner to reason about the interrelationship between these activities. The social layer keeps track of the fact that different subsets of this plan network correspond to different plans – some belonging to the agent and some corresponding to (what the agent believes to be) plans of other agents. The social layer also computes a variety of high-level relations between plans. Plans can contain threats and the plans of one agent can introduce threats or be threatened by the plans of another agent (such relations are computed using the basic plan-evaluation routines provided by standard planning systems). Plans of one agent can also be relevant to other agents (as computed by the plan-relevance criteria proposed by desJardins and Wolverton, 1998). Plans may be interdependent in the sense that one depends on effects produced by another.

Emotional State: The social layer incorporates a model of emotional reasoning, Emile, that derives an emotional state from syntactic properties of an agent's plans in memory [1]. Emile incorporates a view of emotions as a form of plan evaluation, relating events to an agent's current goals (c.f., [4]). Emile computes an agent's overall state, tracks emotions arising from a specific plan, and makes inferences about the emotional state of other agents (given an understanding of their goals and plans). Emotional state is represented as a real-valued vector representing the intensities of different emotional states (Fear, Joy, etc.) and Emile dynamically modifies this state based on the current world situation and the state of plans in memory.

Static State: Static social state components describe features of an agent that are invariant in the course of a simulation. These components can be arbitrary and act simply as conditions to be tested by the social control program. One can manipulate an agent's top level goals, its social status, its etiquette (its sensitivity to certain social cues), its independence (is it willing to construct plans that depend on the activities of other agents), and characteristics of its relationship with other agents (friendly, adversarial, rude, deferential, etc.).

3.2 Control Primitives

Control primitives are social-level actions and consist of communicative and plan-control primitives.

Communicative Primitives: The social layer defines a set of speech acts that an agent may use to communicate with other agents. As they are defined at the meta-level, they can operate on plans only as an atomic structure and cannot make reference to components of a plan (although one has the option of breaking a plan into explicit sub-plans). Some speech acts serve to communicate plans (one can INFORM another agent of one plans, REQUEST that they accept some plan of activity, etc.). Other speech acts serve to change the state of some previously communicated plan (one can state that some plan is under revision, that a plan is acceptable, that it should be forgotten, etc.).

Planning Primitives: Planning primitives alter base-level planning behavior. Classical planning algorithms can be viewed as a sequential decision process: critiquing routines identify problems with the current plan and propose a set of changes that resolve at least one of these problems (e.g. add an action); a change is applied and the process continues. Planning primitives act by constraining the set of viable changes. Recall that from the perspective of the planning algorithm, all activities are represented in a single task network (whether they belong to the agent or represent the activities of other entities). One set of planning primitives allows one to create and manipulate plan objects. Plans can be created and destroyed, and they can be populated with new goals and with activities communicated by other agents. Another set of planning primitives determines whether the planning algorithm can modify the activities in one of these plan objects. One can make a plan modifiable, allowing the planner to fix any flaws with that plan, or one can freeze its current state (as when adopting a commitment to a certain course of action). One can also modify the execution status of the plan, enabling or disabling the execution of actions within it. Finally, another set of planning primitives alters the way the planner handles interactions between plans and thereby implements the idea of a social stance. For example, what happens when Steve detects that his plan conflicts with Jack's. He has several options. He could adopt a rude stance towards Jack, running to grab the keys before Jack gets a chance to take the car. This essentially corresponds to a strategy where the planner resolves any threats that Jack introduces into Steve's plans, but ignores any threats that Steve introduces into Jack's. Alternatively, Steve could take a meek stance, finding some other ways to get to the beach or simply staying home. This corresponds to a strategy where the planner treats Jack's plans as immutable, resolves any threats to Jack's plans, and tries to work around any threats that Jack introduces into Steve's plans. Steve could be helpful, adding activities to his plan that ensures that Jack gets to the market. Or he could be authoritative, demanding that Jack drive him to the beach (by inserting activities into Jack's plans). These stances are all implemented as search control, limiting certain of a planner's threat resolution options. The following are two paraphrased examples of rules that make up Steve and Jack's social control program. The current implementation has about thirty such rules:

```
Social-Rule: plan-for-goal
IF I have a top-level goal, ?goal, ?p THEN
       Do-Gesture(Thinking)
       Say(to-self, ``I want to ?predicate'')
       ?plan = create-new-plan()
       populate-plan(?plan, ?goal)
       enable-modification(?plan)

Social-Rule: you-cause-problems-for-me
IF my plan, ?plan, is threatened by your plan
   I don't have an obligation to revise my plan
   you don't have an obligation to revise your plan
   you don't know my plan THEN
        Say(?you, ``Wait a second, our plans conflict'')
        SpeechAct(INFORM_PROB, ?plan, ?you)
```

4. Summary

Socially situated planning provides one mechanism for improving the social awareness of agents. Obviously this work is in the preliminary stages and many of the limitation and the relationship to other work could not be addressed in such a short chapter. The chief limitation, of course, is the strong commitment to defining social reasoning solely at the meta-level, which restricts the subtlety of social behavior. Nonetheless, our experience in some real-world military simulation applications suggest that the approach, even in its preliminary state, is adequate to model some social interactions, and certainly extends the state-of-the art found in traditional training simulation systems.

Acknowledgments

This research was funded by the Army Research Institute under contract TAPC-ARI-BR

References

[1] J. Gratch. Emile: Marshalling passions in training and education. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 325–332, New York, 2000. ACM Press.

[2] J. Gratch and R. Hill. Continous planning and collaboration for command and control in joint synthetic battlespaces. In *Proceedings of the 8th Conference on Computer Generated Forces and Behavioral Representation*, Orlando, FL, 1999.

[3] B. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

[4] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[5] R. W. Pew and A. S. Mavor, editors. *Modeling Human and Organizational Behavior*. National Academy Press, Washington D.C., 1998.

[6] J. Rickel and L. Johnson. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13:343–382, 1999.

Chapter 23

DESIGNING FOR INTERACTION

Creating and Evaluating an Empathic Ambience in Computer Integrated Learning Environments

Bridget Cooper and Paul Brna
Leeds University

Abstract Central to communication and understanding is the quality of empathy, which enables people to accept, be open to and understand the perspectives of others, whilst simultaneously developing their own perspective. Empathy supports and enables interaction, and creates the climate for both affective and cognitive support. An empathic approach can also be embedded in the design of computer-based learning environments by involving the users at every stage and by supporting a variety of natural human interaction. This chapter explains how this was carried out in a 'classroom of the future'.

1. Introduction

Central to education are the construction of knowledge and the gaining of insights from multiple perspectives. Fundamental to both these aims is the ability to communicate, without which the fusion of widely varying perspectives, which is at the heart of creativity and new understanding cannot even begin to be realised. Complex communication is the corner stone of what it is to be human, to envisage, to abstract, to reflect upon and interact with our environment and its inhabitants. Central to communication and understanding is the quality of empathy, which enables people to accept, be open to and understand the perspectives of others, whilst simultaneously developing their own perspective. Empathy supports and enables interaction, and creates the climate for both affective and cognitive support. An empathic approach can also be embedded in the design of computer-based learning environments by involving the users at every stage and by supporting a variety of natural human interaction. This theoretical understanding underpinned a European project NIMIS (Networked Interactive Media in Schools) which envisaged a classroom of the future through the development of intuitive hardware and software designed to develop collaborative skills, perspective taking and literacy skills in pupils from 5-8years. This chapter focuses on the UK section of the project. It also builds on the findings of a research project, which looked at the role of empathy in teacher/pupil relationships. The chapter argues that the quality of human communication and interaction is central to the ambience in learning environments and that high quality empathic design can support that ambience.

2. Perspective

The work is situated within a broad framework of educational and technological research into school ethos [14] and teaching and learning research into empathy [1, 17], self-esteem [16], communication and dialogue [20, 3], prior learning [2] and effective teaching and learning. Additionally this chapter considers some recent empirical research into empathy in teaching and learning [5]. ICT can be liberating but needs careful design and evaluation in the human contexts in which it will be used, allowing humans and ICT to work creatively together, maximising the strengths on both sides. One area where technology can contribute enormously to the problem of time and resources is by the provision of one to one or small group support [10]. Pupils motivated by the use of technology, by its practical, flexible, and often, exciting potential, are able to take greater control of their learning. Teachers are freed up by this to take a more facilitative role, devolving responsibility for learning to pupils, as found in Machado and Paiva's work in a Portuguese school working with ICT to promote learning through drama for children aged around 8 years old [13]. This places them in a more empathic position with pupils, with less need for traditional teacher domination and control and in a better position to work with and understand individuals, thereby modelling an empathic approach which pupils are likely to imitate. These closer, more equal, more human, relationships are likely to promote better assessment practices and through them improve learning [8]. However the quality and effectiveness of the technology and the training of teachers in its use are important factors. An irate teacher, struggling with temperamental computers, and/or inappropriate software will find it difficult to model empathy to anyone or encourage it others — so teacher involvement in design and training was built into the NIMIS project from the outset. Intelligent software, which utilises knowledge of teaching and learning and attends to the varying needs of all learners, at the appropriate moment, is necessary for optimum learning to take place. Software which appreciates the significance of the affective elements in teaching and learning and that these are inseparable from the cognitive aspects [6, 11] is more likely to engage the learner. Intelligent agents who create a positive atmosphere through affirmation and appropriate feedback to develop language and narrative skills reinforce the positive environment created by the teacher. A positive, nurturing, and enabling atmosphere, which supports all children, provides the model for their own personal development and supports their relationships and empathy with others. We believe that flexible classrooms designed to meet children's needs, to encourage a wide range of interaction and collaboration, to enable the co-construction of ideas, presentation of ideas and subsequent reflection, can help to support and nurture both the emotional, social and intellectual development of children. The NIMIS classroom aimed to provide a variety of opportunities for presentation, interaction and reflection through the provision of a number of different shared workspaces, co-operative layout and also electronic interaction as well as affective support through the development of an empathic agent who would combine the affective and the cognitive support at the computer to supplement the human support available from teacher and peers, thus maximising the interaction.

3. The Significance of Empathy

Developing a rich and sensitive understanding of every child requires considerable empathy on the part of the teacher. Research into empathy and into teaching and learning in the '60s and '70s explored the concept in considerable depth and linked empathy with effective teaching. Aspy (1972) and Rogers (1975), amongst others, highlighted the central nature of this quality not only in teaching but in all caring relationships. Empathy is widely associated with development and learning from intensely personal development during therapy to intellectual, spiritual, creative and also moral development [12]. Teachers are obliged both to discover a pupil's existing skills or understanding in a particular subject area and extend them but in order to do this most effectively they have to know the child as a person, know their confidence levels as well as be aware of their academic understanding. They have to nurture their sense of self and support their academic success, which can also further develop their sense of self. They may also develop their students awareness of other people, through simultaneously valuing them and opening their eyes to other attitudes and understandings very different from their own. Empathy and the interaction it involves therefore are central to developing understanding of subjects, skills and of other human beings. The study on empathy in teacher/pupil relationships is UK based and involved recorded interviews with pupils and teachers and observations of them at work in the classroom using tape-recorders and field notes [5]. Later, it involved secondary and primary schoolteachers of different genders and subject specialisms and degrees of responsibility who were especially selected for their empathic approach to teaching and learning. Analysis followed grounded theory methodology [19]. These teachers understood empathy to be essential to their teaching. They described how it created trust, nurtured feelings of security, built confidence and enabled two-way communication. The teacher had to learn about the child as much as the child had to learn about the subject. Empathy was central to high quality, effective teaching and learning, enabling greater understanding, better assessment, better academic and emotional support and consequently more appropriate teaching provision and more appropriate differentiation. Empathy equalises relationships, valuing children's contributions and understanding and allowing them more control over their learning. Empathy enables the right support, to be given at the right time, ensuring better scaffolding.

4. Methodology

Empathy built into the methodology of project design involves valuing existing knowledge and understanding and recognising best practice. Sensitive system design reflects this by involving all participants from the outset [7]. Rapid prototyping in real situations with continual feedback from pupils and teachers, coupled with theoretical reflection on the outcomes from a more detached perspective, is more likely to ensure appropriate and responsive learning systems. The complexity of evaluating the use of technology in education is well documented and there are both established and evolving schemes to support and illuminate the process [9, 15]. We chose to adapt a methodology, first developed by Carroll and Rosson [4]. This methodology is one of several participatory design approaches, and is organised around the identification and exploration of key scenarios. Additionally, it uses a form of design rationale known as claims analysis. We extended the claims concept to incorporate the pedagogical intentions underlying the design, and called this pedagogical claims analysis. With this method of evaluation each aspect of the design process is linked to some possible pedagogic outcome, which raises possible issues and suggests how each particular claim might be checked. During the initial collection of pedagogical claims, children and teachers were engaged in low-technology design, [18]. The evaluation has both formative and summative aspects. The initial claims are revised and validated throughout the formative prototyping phase. The claims help the understanding of the design process and making design decisions explicit. Generating many claims in a complex project helps determine priorities. The way the classroom functioned was observed before the technology was introduced and teachers met to discuss and share their ways of working and teaching methodology for literacy. They developed typical scenarios of existing ways of working and then envisaged ways in which the new technology might support/enhance the functioning of their existing classrooms. Hence teachers helped to design both the classroom layout and contributed their ideas and understanding to the software development before participating regularly in the ongoing prototyping and evaluation procedures. Design and evaluation was a continuous process, the one naturally leading into the other with the expectancy and the result that the classroom would meet the needs of the teachers and children very successfully. The NIMIS classroom eventually included a small network of computers, invisible but for the small touch sensitive screens (WACOM tablets) and pens which can lie flat on the desk and are thoughtfully laid out to encourage collaboration and group work around a small table. There is also a large touch sensitive screen especially modified for use by small children, which they can operate with their fingers. The software, designed to encourage collaboration, literacy development and story writing is distributed allowing pupils to create stories together and exchange ideas and reflections across the network, whilst at the same time communicating by natural means. The software was designed to meet a range of attainment levels in order to empower learners and support diversity and development. Creating shared stories or reflecting on them and supporting each other in the writing allows children to understand different perspectives which also helps to develop empathy. Features to encourage the creation of stories with different perspectives is also embedded in the software structure. This takes the form of thought and speech bubbles, real and fantastic situations and characters and word banks with speech synthesis, which link the known to the unknown in both sound, picture and text. An empathic agent modelled on the helpful support of teachers in one to one situations from the study on empathy is embedded in the software and can also add to the positive ambience of the classroom. The agent can offer affirmation followed by support based on knowledge of the child, the creation of stories and the features, which the child has chosen. To get a holistic view of the classroom and to evaluate the human and technological aspects of the classroom in combination we used a variety of data collection methods. These included video recordings of lessons, teacher diaries and interviews, children's interviews, researchers

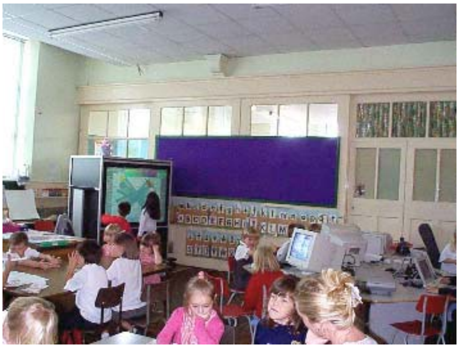


Figure 23.1. The classroom nearing the end of its development

field notes, evaluation of the stories children produced, computer logs of the activity on the computers, as well as National Curriculum tests, reading tests and some limited comparison with another year one class working with the same curriculum. The subsequent analysis looked closely at the quality of the human interactions in this classroom as well as the computer interactions and the stories produced.

5. Outcomes

We present a very brief summary of relevant evaluation data below recorded over the academic year from September 1999 to July 2000. There were 23 children in this year 1 class (5 & 6 year olds). In particular we examine the issues relating to ambience and interaction. The large screen and the network around the octagonal table were used daily for up to five hours/day. The technology was thoroughly integrated into daily aspects of teaching and learning. The enthusiasm, engagement in and enjoyment of the NIMIS classroom continued throughout the year. Children and teachers were highly complimentary about the facilities it provided. A typical child's comment when asked about having the classroom for a year was "because its really nice and people love it. They always want to play with it all the time... it makes me feel happy and feels nice."

In the final interviews the teachers remained very pleased with the whole classroom and described it as "wonderful", "a perfect world" and "I wouldn't be without it". Reflecting on the children's attitude they echoed the children's feelings, "they love it... and at the end of the day they love to go there (on the computers) they still have the same amount of enthusiasm (as at the start)". The teachers explained how the classroom helped with their teaching, because of the flexibility of the large screen and table of small WACOMs and a range of software they could integrate the computers very easily into their teaching in a very natural way. The teachers were able to engage the whole class at one moment through the clarity and versatility of the large screen and then use the network around the octagonal table to motivate and support low attaining pupils. Emotional excitement on the part of the teacher also transmits itself to the children and draws them into the learning, increasing the interaction and engagement. A strong sense emerges from the final interviews of both teachers and children that a good helping atmosphere was present in the class. There is evidence to show that collaborative and helping behaviours were encouraged and that children had opportunities to gain confidence in front of each other and by explaining things to each other. In this sense the empathic and interactive ambience that we had hoped to create did appear to emerge in this classroom. Levels of engagement in tasks and interactions were over twice as high when children were using the computers as they were before their introduction. From the point of view of affect the children were extremely positive about the NIMIS classroom. Enjoying what they do helps to motivate them, makes them feel confident and independent but also co-operative and motivates them to learn. One teacher explained how they are quite happy to do additional Maths and English in the afternoons on the computers. The teachers too felt very pleased and happy with the classroom, which is also likely to influence the children, since teachers establish the general climate in the classroom. The analysis of the interactions showed the interaction with adults to be generally of a higher quality but less timely (due to teacher/pupil ratio) and the interactions with peers to be usually of a lower quality but more timely (due to availability). These different forms complement each other in a busy classroom but an empathic agent which could provide both affective and cognitive support in the story-writing software could contribute to both the quality and timeliness of the interactions as well as modelling high quality interaction to young children to improve their ability to support others.

6. Conclusion

The recent re-emphasis on of the importance of the emotions is perhaps one of the most significant developments in understanding the learning process. Brain research increasingly suggests that the cognitive and the affective are inextricably linked and perhaps only holistic approaches and evaluations can really begin to understand the nature of high quality learning environments. The NIMIS classroom and software was designed with a belief that all these factors work together to make successful learning. Computer integrated classrooms can maximise the strengths of both people and computers, supporting interaction of many different kinds and combinations. The success of both classroom and software has provided considerable justification for our thinking and methodology in the project. Our future aim is to develop the empathic agent to further improve the ambience of the classroom at the level of one-to one personal interaction and to complement the teacher and peer support. The enthusiastic responses of both teachers and children, coupled with strong evidence in the video analysis of very high levels of engagement and greater opportunities for collaboration suggest that such a classroom presents a very positive model for humans and computers working effectively together. The aim of smooth and natural interaction between the human and the digital does seem to have occurred. The holistic view which took into account affect and relationships, communication and interaction as being central to the learning process has contributed to creating a positive classroom climate in which children are motivated, confident and mutually helpful.

Acknowledgments

This work was supported by the EU funded Networked Interactive Media in Classrooms (NIMIS) project no 29301.

References

[1] D. Aspy. *Towards a Technology for Humanising Education.* Research Press, Champaign Illinois, 1972.

[2] N. Bennet and E. Dunne. How children learn — Implications for practice. In B. Moon and A. Shelton-Mayes, editors, *Teaching and Learning in the Secondary School*, chapter 6, pages 50–56. Routledge, London, 1994.

[3] J. Bruner. *Child's talk: Learning to use language*. Oxford University Press, Oxford, 1983.

[4] J. Carroll and M. Rosson. Getting around the task-artifact cycle: How to make claims and design by scenario. *ACM Transactions on Information Systems*, 10:181–212, 1992.

[5] B. Cooper. Exploring moral values — knowing me knowing you - aha — rediscovering the personal in education. In R. Best, editor, *Perspectives on Personal, Social, Moral and Spiritual Education*. Cassell, London, 2000.

[6] Antonio R. Damasio. *Descartes' Error: Emotion, Reason and the human brain*. Macmillan, London, 1994.

[7] A. Druin. The design of children's technology. *Morgan-Kauffman, USA*, 1999.

[8] M. Drummond. *Assessing Children^{1}s Learning*. David Fulton, UK, 1993.

[9] S. C. Ehrmann. Studying teaching, learning and technology: A tool kit from the flashlight program. *Active Learning*, 9:36–39, 1998.

[10] M. Elsom-Cook. Guided discovery tutoring and bounded user modelling. In J. Self, editor, *Artificial Intelligence and Human Learning*. Chapman and Hall, London, 1988.

[11] D. Goleman. *Emotional Intelligence*. Bloomsbury, 1995.

[12] M. Hoffman. Moral development. In P. Mussen, editor, *Carmichael's Manual of Child Psychology*. Wiley, New York, 1970.

[13] I. Machado and A. Paiva. Me, my character and the others. This volume.

[14] Department of Education and Science. Discipline in schools. *HMSO, London*, 1989.

[15] M. Oliver and G. Conole. Evaluating communication and information technologies: A toolkit for practitioners. *Active Learning*, 8:3–8, 1998.

[16] W. Purkey. *Self-concept and School Achievement*. Prentice-Hall, 1970.

[17] C. Rogers. Empathic: An unappreciated way of being. *Counselling Psychologist*, 5(2):2–10, 1975.

[18] M. Scaife, Y. Rogers, F. Aldrich, and M. Davies. Designing for or designing with? Informant design for interactive learning environments. In *CHI'97: Proceedings of Human Factors in Computing Systems*, pages 343–350, New York, 1997. ACM.

[19] A. Strauss and J. Corbin. *Basics of Qualitative Research: Grounded theory procedures and techniques*. Sage, Newbury Park, CA, 1990.

[20] L. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA, 1978.

Chapter 24

ME, MY CHARACTER AND THE OTHERS

Isabel Machado[1] and Ana Paiva[2]
[1] *ISCTE, INESC, CBLU - University of Leeds*
[2] *IST-Technical University of Lisbon, INESC*

Abstract Dramatic games can develop communication skills, instill confidence in children and encourage teamwork. Inspired by these games we have developed a collaborative virtual story-creation environment called *Teatrix*. *Teatrix* is an innovative application, where children may build their own "virtual stories". Going further than merely providing the children with the means to act their stories, *Teatrix* also offers them the possibility to reflect upon their own actions during story creation time. This chapter explains how children can create their stories in *Teatrix*, by controlling their synthetic characters in a 3D world.

1. Introduction

Drama plays an important role in children cognitive development. Dramatic games can develop communication skills, instill confidence in children and encourage teamwork [8]. Based on these findings, and focusing on the role of collaboration in dramatic games we designed a system called (*Teatrix*), which was essentially inspired by many of the activities that exist in children's dramatic games.

With *Teatrix* children can create virtual plays by controlling the characters in 3D worlds. Each child controls her character, which means that she has to: (1) take into account her character's role in the story; (2) reflect upon why her character has taken a certain action, i.e., taking her character's perspective and respond for its behaviour; (3) consider the other children's opinions about the story, in order to contribute meaningfully for their collaborative task.

In this chapter we will describe the experiences we had with dramatic games, which led to some design decisions in *Teatrix*. These decisions aim at fostering collaboration and the emergence of a common story by a group of children. We will argue that such emergence needs some high level discussion, which is not achieved by the simple control of synthetic characters at a simple action level. Then we will provide some discussion on how to achieve such high level discussions and present some results already obtained.

2. Context

To better inform the design of *Teatrix*, we performed a set of studies in a Portuguese school called "*O Nosso Sonho*". This school's pedagogical approach follows the argument that each child must be free to choose her own daily activities[1]. To do so, "*O Nosso Sonho*" has several thematic rooms in which the children can experience different types of activities. One of such rooms is dedicated to dramatic games. There, we conducted a set of observations of children's story performances. To test different factors influencing the resulting performances, we selected two groups of children of different ages (one group of 4 to 6 years old and another of 7 to 9 years old). With these two groups, we wanted to observe not only how the children engaged in the story performance but also what types of collaboration existed during acting. The stories performed by both groups were the following fairy tales: "*The three little pigs*", "*Cinderella*" and "*Hansel and Gretel*".

The analysis of the observations was not trivial as interactions in dramatic games occurred at different levels, and in parallel between different participants. Nevertheless, from the data collected (video recording and notes taken before, during and after the performances) we were able to identify several factors that influenced, not only the acting, but also the collaboration threads that emerged from these performances. These factors were:

- *Age of the children* - for the younger group the dramatic games were important not for resulting "play", but mostly because of the interaction opportunity that emerged from the activity. Young children had difficulty to *stay in character* and the story itself was not achieved in the majority of cases. In fact, the use of dramatic games at this age group aims at promoting children's interactions and also the experiencing of a multitude of different situations that can help such children to deal with their inner fears [1]. Collaboration, inside or outside the story, occurred rather occasionally and usually in situations where two friends interacted with each other. Differently, the older group was more prone to *stay in character* and the children's interactions were mostly originated by the plot itself. In this group, the collaboration occurred not only inside the story between characters but also as a way to coordinate their positions, turns, props, etc.

- *Inter-relationships* - during the performances children, specially the younger ones, had the tendency to bring their daily relationships into the story, and so it was common to see two friends in close interaction inside the story (and even forget their roles).

- *Teacher intervention* - the teacher was always present in all the performances. In the case of the younger group, the teacher had much more active interventions and control. Usually, the teacher played the role of *director* and *narrator* of the story. Differently, the children in the older group were able to play these roles by themselves (often, one of the characters would spontaneously turn into a kind of director), leading the teacher to the position of a spectator.

- *Presence of audience* - we observed performances with and without audience. On the one hand the audience provided a motivational factor and influenced the children's performances (they wanted to show the others how well they acted). On the other hand the audience played the role of a distracting and sometimes perturbing factor.

As a result of the observations, we decided to focus our research and development in the age group of 7 to 9. Further, from the observations we were able to identify some features of children's dramatic games, which served as functional elements in the design of *Teatrix*. Some of these features were:

- *Phases*- there are several phases in the dramatic games. The first phase- the *preparation* includes: story selection and discussion, selection of the characters and the choice of the actors and props. The second phase is the *acting* itself. The final phase is a *discussion* phase about the performance done.

- *Action and use of props*- children selected from their classroom several different objects that they used for the acting, ascribing them different meanings.

- *Interaction between the children*- we were able to distinguish two types of interactions between children: (1) "performance level interactions" when children interact through their characters by their actions and sentences; and (2) "co-ordination interactions", when children provide signals to the others, give orders, make demands or simply inform the others about issues related with the ongoing play. Note that this type of coordination is done during the performance, as the play develops.

3. Application: *Teatrix*

Taking the results of our observations we developed *Teatrix* [2] which aims at providing effective support for children developing their notions of narrative through the dramatisation of different situations (reflecting a participatory design approach taken in the NIMIS project - see [3]). Inspired by the ritual found in dramatic games of the school, *Teatrix* is divided in three phases. In the first phase, one child (often helped by the others) is responsible for the preparation of the story. This preparation includes the choice of the theme, the cast, props and scenes for the future play (see Figure 24.1).

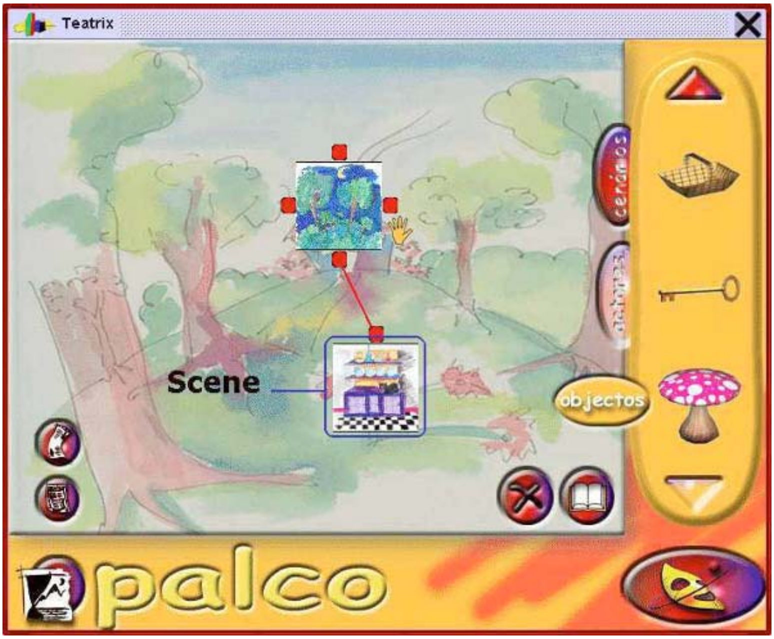


Figure 24.1. *Teatrix*: *Backstage* Option

The second phase provides the children with the possibility to initiate one story and to start the acting (on *stage* performance). This option is accomplished in a collaborative 3D world (see Figure 24.2) where each participating child controls one character. So, at the beginning of the acting phase, each child has to choose her character to control. The story creation only evolves if the children try to achieve a common story, collaboratively. However, the way the story will evolve at acting time is completely open to the goals of the group of children engaged in its creation.

Once the play is finished, children obtain a product- a "*film*"-like object, which they can exhibit as a proof of their collaborative work. This "*film*" is then used for the next phase. In this last phase, children can be the audience of their own performances and watch their previous stories. The story can be seen from different perspectives (focusing the camera on different actors), which can provide the children with the possibility to discuss each others' performances or even write (using a special tool also developed within the NIMIS project) different stories based on the same performance.

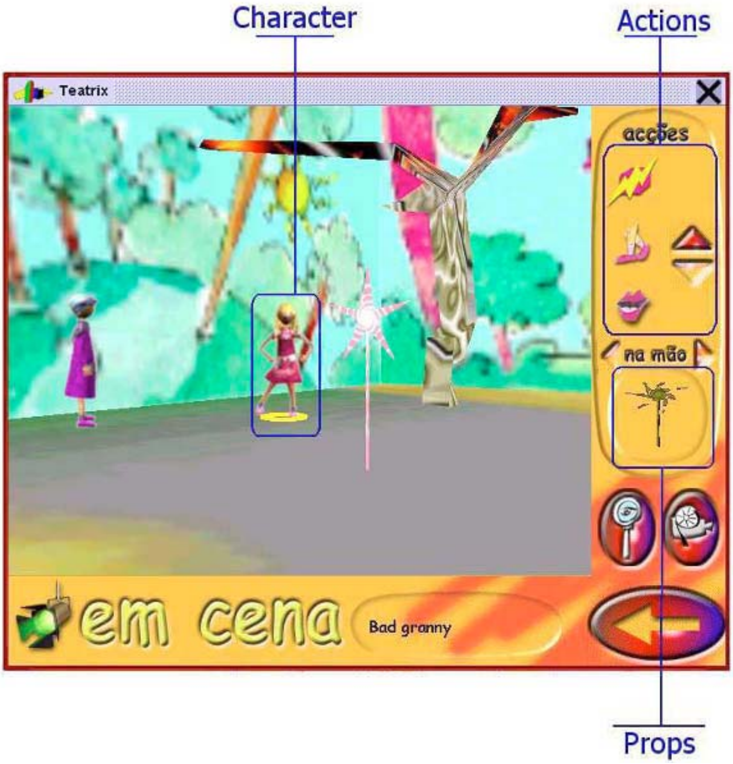


Figure 24.2. *Teatrix*: *On Stage* Option

4. Controlling the Characters in *Teatrix*

In the acting phase each child has a character to control. Characters in *Teatrix* are not simply controlled animations in a 3D world. They also have roles (villain, hero, etc), which define their function in the story (for example, a villain must "harm the hero", a hero must "rescue the victim", etc). So, a character must be seen as a conjunction of two different concepts: the *actor* and the *role*. An *actor* is the physical representation or appearance of a character. In *Teatrix* we provide a cast that contains a witch, a boy, a girl, a fair, a wolf and an old lady. More actors can be added provided that the appropriate animations are built into the system. As for the roles, we relied on the analysis done by Propp [7] on folktales, which led us to the following roles:

hero/heroine, villain, donor, magician, loved one and family. In *Teatrix*, roles define and establish the function of the agent (character) in the narrative, by means of the specification of particular actions and goals for that agent (e.g., a villain character has as one of its goals harming the hero). With these well defined sets of roles and actors, we aimed at providing the children with a set of varied characters who are interesting, identifiable and fun and at the same time that have the means to develop and grow throughout the story creation process [8].

Furthermore, the characters were developed in such a way that they can act autonomously in the story, if not controlled by a child. This is simplified by the fact that the system must try to guarantee that the character follows the role that was assigned to it. This means that a role has associated a set of goals that the system will try to achieve (see [6] and [5] for a more detailed description of the agent's architecture).

To control the characters *Teatrix* provides the children with a set of actions which they can select at acting time (see Figure 24.3). These actions are associated not only with the character performing it but also with the props that the character owns at each instant (see Figure 24.3). In our research, we have embedded inside the objects the necessary knowledge of what effect they will have in the environment.

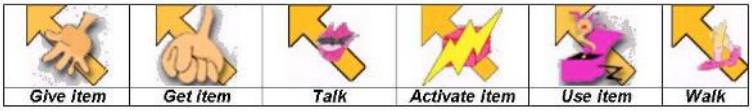


Figure 24.3. Actions for controlling the characters

This defined set of actions provides the children with motion control (for example: each child can move her character along the scene by using the move action) and a type of behaviour control, achieved through assignment of a role to the characters and with the use of the props.

5. Communication and Reflection in *Teatrix*

Taking into account the types of interactions observed in the dramatic games of the school, we tried to provide some mechanisms for communication embedded in *Teatrix*. Basically, children can communicate through their characters in two ways: 1) by using their objects on another character (for example one child may use her character's stick on another character to harm it); or 2) by speaking to another character using the "talk" action.

These two ways of communication were broadly used by children in the version installed in the NIMIS classroom of the school "*O Nosso Sonho*".

However, after some sessions with this first version children started to demand more understanding on what was really happening in the story. Also, they would ask aloud their peers about what was happening. In general, they demanded some higher degree of control over their characters. They wanted to better understand what the characters meant when performing a certain behaviour. To respond to their demands we introduced a new type of control over the characters.

The idea was to offer the children with the possibility to reflect upon their characters' behaviours at story creation time and control that behaviour. This meta level of control is implemented as a tool called the "Hot-Seating", which gives the children the possibility to freeze the *story time*, put themselves into their characters' shoes and explain the character's behaviours [2]. When a child enters the "Hot-Seating" she is asked to answer a set of questions concerning the behaviour of the character. Such justifications are then shared with all the other children in the same virtual play.

These *reflection moments* may happen at the child's demand or when the application detects that a character is not *in character* (see [4] for further details). With this tool we aimed at providing the children with more information about the story, which, we believed, would lead to a richer type of collaboration.

We have installed a new version of *Teatrix* in the school and so far the results are quite positive. In spite of the fact that in the first two weeks children tended to ignore the reflection moments (even if triggered by the system), in the last couple of weeks they started to use more often the reflection tool and to justify their character's behaviour. So the "Hot-Seating" was easily understood. However, we still haven't found any significant results that establish a relation between the presence of the "Hot-Seating" and the quality of collaboration established between peers in the same virtual play.

6. Final Remarks

This chapter provided an overview of some of the design decisions taken in the construction of *Teatrix*, a collaborative virtual environment for story creation by young children. We described results of the observations made with children performing dramatic games, and, based on these observations we introduced a new approach for character control and communication in *Teatrix*.

Notes

[1] Note that “*O Nosso Sonho*” is not a curricular school.

[2] *Teatrix* is an application that was developed under the Networked Interactive Media In Schools (NIMIS) project, a EU-funded project (n. 29301) under the Experimental School Environments (ESE) program.

References

[1] B. Bettelheim. *The uses of enchantment: the meaning and importance of fairy tales*. Harmondsworth, Penguin, 1978.

[2] G. Bolton. *Acting in classroom drama : a critical analysis*. Stoke-on-Trent : Trentham, 1998.

[3] B. Cooper and P. Brna. Designing for interaction – Creating and evaluating an empathic ambience in computer integrated learning environments. This volume.

[4] I. Machado, P. Brna, and A. Paiva. Learning by playing: supporting and guiding story-creation activities. In *Proceedings of International Conference on Artificial Intelligence in Education*, San Antonio, USA, 2001. IO Press.

[5] A. Paiva, I. Machado, and R. Prada. The child behind the character. *IEEE Journal of Systems, Man, and Cybernetics, Part A*, 31(5): 361–368, 2001.

[6] A. Paiva, I. Machado, and R. Prada. Heroes, villains, magicians,...:dramatis personae in a virtual story creation environment (forthcoming). In *Proceedings of the Intelligent User Interfaces Conference*. ACM Press, 2001.

[7] V. Propp. *Morphology of the folktale*. Austin: University of Texas Press, 1968.

[8] D. Wood and J. Grant. *Theatre for Children - A Guide to Writing, Adapting, Directing and Acting*. Faber and Faber, 1997.

Chapter 25

FROM PETS TO STORYROOMS

Constructive Storytelling Systems Designed with Children, for Children

Jaime Montemayor, Allison Druin, and James Hendler
University of Maryland Institute for Advanced Computer Studies

Abstract Working with children as our design partners, our intergenerational design team at the University of Maryland has been developing both new design methodologies and new storytelling technology for children. In this chapter, we focus on two results of our efforts: PETS, a robotic storyteller, and Storykit, a construction kit of low-tech and high-tech components for children to build physical interactive storytelling environments.

1. Introduction

Since 1998 our interdisciplinary and intergenerational design team at the University of Maryland has been developing new technology for children, with children. Our team blends children (7-11 years old) and adults, from disciplines as diverse as engineering, education, computer science, and art. In large part, because of our child design partners, we have come to focus our work on the development of technology that encourages storytelling for elementary school-aged children, and most recently for kindergarteners.

Because storytelling is inherently constructive, and since children explore new ideas and feeling through stories ([6], [15], [20]), the resulting products of our design team have been kits that enable children to create their own stories. Our research projects have evolved from PETS, an emotional storytelling robot [10], to a StoryKit that enables children to build physical and interactive story environments [1]. In this chapter we will first briefly describe *cooperative inquiry* ([7], [9]), our design methodology for including children as design partners. We then use the PETS and StoryKit projects to demonstrate how storytelling technologies can enhance creativity, collaboration, and social interactions among elementary school-aged children.

2. Our Design Approach: Cooperative Inquiry

While many participatory design techniques exist for including adult users into the design process, these same approaches are not always appropriate for children. *Cooperative inquiry* is a collection of techniques adapted and modified from existing methodologies to suit the special needs of an intergenerational design team ([7], [8], [9]). Its three components are: contextual inquiry, participatory design, and technology immersion.

Contextual inquiry, based on the work of Beyer and Holtzblatt [2], is a technique for researchers to collect data in the users' own environments. Rather than a single text-based note-taking method, we suggest adult and child researchers each record their observations with different methods. So, adults may record their observations with text, while children draw cartoon-like pictures to describe their observations. (See [7] for specific note-taking techniques.)

In our *participatory design* sessions, we construct low-fidelity prototypes from material such as crayons, cardboard boxes, LEGO blocks, and fabric, because they are easy to use by both adults and children. These constructed artifacts become the bridge for discussions between adults and children.

While adults may have access to technologies throughout their workday and at home, the same is less common for children. Therefore, we have found *technology immersion* to be an important time for children to use technologies as much or as little as they choose.

3. Related Work

Researchers over the past few decades, recognizing both children's innate abilities and the potential afforded by new technologies, began designing new computational devices that encourage self-learning ([21], [23]). Some successful systems use robots to engage children in the discovery of scientific and mathematical principles (e.g., [12], [16], [21]). More recently, robotic storytellers have also been explored and developed for children, including, SAGE [26] and Microsoft Actimate Barney [25]. Other robots, such as KISMET [5] and Sony's AIBO [13], allow researchers to study social contexts such as behaviors and emotions. Our PETS robot conveys emotions in stories by performing gestures that elicit sympathetic responses from its audience.

While physical interactive environments have traditionally offered entertainment (e.g., DisneyQuest), education in the sciences (e.g., [24]), and self-expression (e.g. art museums), researchers have recently begun exploring them as a medium for storytelling. Unlike most systems that are constructed and programmed by technologists for the novice users (e.g., [11], [3]), props and interactions inside StoryRooms [1] are constructed by children for themselves.

4. A Storytelling Robot

PETS, a "Personal Electronic Teller of Stories," is a robotic storytelling system for elementary school age children ([10], [19]). The PETS kit contains a box of fuzzy stuffed animal parts and an authoring application on a personal computer (figure 25.1). Children can build a robotic animal pet by connecting animal parts such as torsos, heads, paws, ears, and wings. Children can also write and tell stories using the *My PETS* software. Just as the robotic animal is constructed from discrete components, *My PETS* is also constructive. This application enables children to create emotions that PETS can act out, draw emotive facial expressions, give their robotic companion a name, and compile a library of their own stories and story starters. Each emotion that the robot performs is represented by a sequence of physical movements that conveys a specific feeling to the audience. Our child designers defined six basic emotions: happy, sad, lonely, loving, scared, and angry. They were chosen because the actions that represent these emotions are sufficiently different from each other that the audience would not confuse one from another. To express loneliness, the robot lowers its arms and looks left and right, as if looking for a friend. When the robot is happy, it waves its arms quickly, turns its head left and right, and spins around. When the robot is sad, it lowers its arms and head, and moves forward at a slow, deliberate pace.

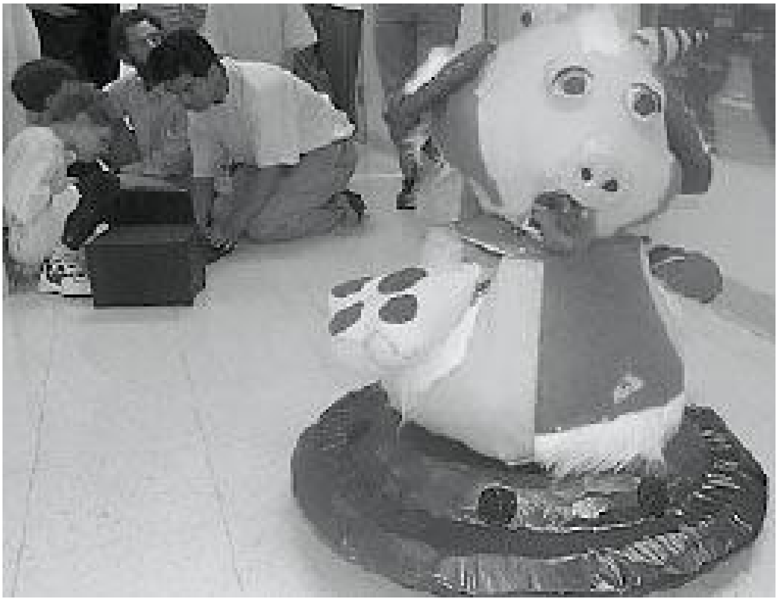


Figure 25.1. Children and adults play with PETS at the 1999 HCIL Open House.

Children write stories using *My PETS*. A simple parsing function detects words that match its list of emotional keys. As *My PETS* recites the story (using text-to-speech), and recognizes an emotion, it issues the corresponding sequence of motion commands to the robot.

PETS supports the reactive and sequencing layers of a multi-tiered architecture (e.g., [4]). The reactive layer is written in Interactive C for the Handy Board microcontroller [17]. The sequencing layer, written in RealBasic, is embedded into *My Pets*, and runs on a Macintosh Powerbook. The two robotic components communicate with *My Pets* through custom-built RF transceivers. The robot contains two distinct components, the "animal" and the "spaceship." Both are made from polycarbonate sheets and steel posts. Servomotors on the animal controls its mouth, neck, and limbs. The spaceship uses two modified high-torque servomotors to drive independent wheels.

Our current work uses a new version of PETS as a motivational tool for children with disabilities to complete their physical therapy [22].

5. Our Second Project: Storyrooms And Storykits

The transition from storytelling robots to storytelling environments was influenced by the limits of robots as actors. Although a physical robot can be an actor, some story elements are either inconceivable or awkward to express. While the robot can project sadness or happiness, it might have difficulty suggesting that "*it was a dark and stormy night.*"

In the summer of 1999, we began work on a technology that would enable children to construct their own physical interactive environments. The lessons we learned from PETS, such as sequencing physical events to form abstract ideas, formed the foundation of this new research focus. We believed that children can construct their own *StoryRooms* from using parts inside a *StoryKit* [1], and that through interactions within this environment visitors can have a new kind of storytelling experience.

Using a prototype *StoryKit*, we built a *StoryRoom* based on the Dr. Seuss story, "The Sneetches" [14]. This is a story about the Sneetches that lived on a beach. Some had stars on their bellies, while others did not. The star-bellied Sneetches believed they were better than the plain-bellied ones. One day, Mr. Sylvester McMonkey McBean arrived and advertized that his inventions could put a star on any plain bellies for just three dollars a piece. Of course, the plain-bellied Sneetches jumped at this opportunity. The previously "better" Sneetches became upset as there was no way to tell them apart! Not surprisingly, Mr. McBean had another machine that took stars off too. As the Sneetches cycle through both machines, one group wanting to be different, the other wanting to be the same, they squandered all their money. Ultimately they realized that they were all the same, whether or not they had a star on their bellies.

We wanted to express this story through a StoryRoom. In our adaptation, children became the Sneetches by wearing a special box, which has a star-shaped cutout and an embedded microcontroller connected to a lightbulb, on their bellies. We then turned our lab into the Sneetches StoryRoom (figure 25.2) by placing the Star-On, Star-Off, Narrator, Mr. McBean, and Money props. The Star-On and Star-Off were cardboard boxes with colored paper glued over it. On each, we attached a light bulb and a contact sensor. The Narrator and Mr. McBean were applications that recorded, stored, and replayed digitally recorded passages from the story. The Money application controlled a projected image of a pile of money, with the Sneetches on one side, and Mr. McBean on the other side. Finally, the boxes on the children's bellies were the Stars that can turn on and off. To help convince the children that the stars made a difference in their social standings, we added a Toy prop, which responded only to those with stars on their bellies. In effect, interactions with the Toy made the children feel as if they were the Sneetches.

When children initially entered our Sneetches room, the star boxes on some of their bellies lit up, while others did not. Next, the Narrator introduced the story. These children explored the room and discovered the Toy. They also noticed that the Toy lit up only for those who had stars on their bellies, but not for those who did not.

Soon, Mr. McBean introduced himself and told the children about the Star-On machine. When a child without a star on her belly crawled through it, her belly lit up; she heard Mr. McBean thanking her for the three dollars she "paid" him and the "ka-chink" of a cash register; she sensed the Star-On box lit up as she passed through it; finally, she saw that some of the Sneetches' money had moved from their pile over to Mr. McBean's pile. Most importantly, when she went to the Toy, it lit up for her! This story continued, until all the money had been spent, and concluded with some final words from Mr. McBean and the Narrator.

6. Observations

At our 1999 Human Computer Interaction Lab Open House, our child design partners showed PETS to other children. They were eager to type in stories to see what PETS would do. Indeed, they wrote at least half-dozen short stories within half an hour. They also enjoyed changing PETS' facial features. One child even turned PETS into something that could belong in a Picasso painting. We also noticed that children responded to the robot's "emotions" because its actions were similar to what they would have done had they felt the same way.

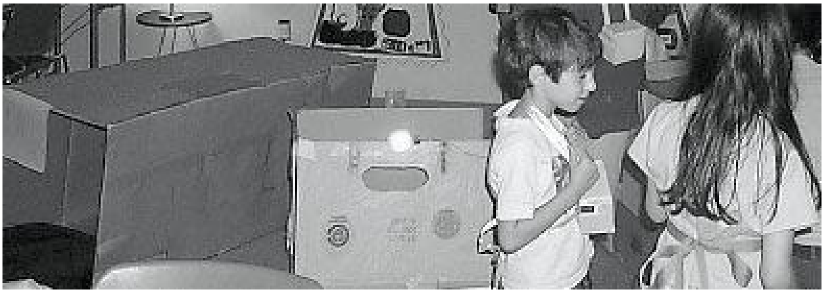


Figure 25.2. Children, with stars on their bellies, experience the Sneetches StoryRoom. The cardboard box on the left is the Star-Off machine. The box in the middle, The Toy, has a light effector attached to it.

Furthermore, stories were more interesting because emotions were more than words on a page, they were also acted out. Indeed, these observations suggest that, at least for our child researchers, perception is sufficient for conveying feelings in stories.

At the end of our summer 1999 design team workshop (an intense 2 week long, 8-hour day experience), we held an open house and invited guests and families to experience our Sneetches StoryRoom. We arranged the visitors into pairs of adult and child designers. They entered the room three pairs at a time. While all the children appeared to enjoy exploring the room and making things happen, their parents did not always understand what was happening. Furthermore, when they activated many things at once, the room became a cacophony, and the story became difficult to follow. We were also pleasantly surprised by their high level of enthusiasm in guiding their guests through the StoryRoom. Not only did these children wanted to build the story, they wanted to share it with others.

Based on observations from our intergenerational collaboration, we created the following guidelines for designing attractive and entertaining storytelling environments for children:

1 Give children the tools to create.

2 Let children feel that they can affect and control the story.

3 Keep interactions simple.

4 Offer ways to help children begin stories.

5 Include hints to help children understand the story.

6 Make the technology physically attractive to children.

Our work continues today on StoryRooms. We are currently developing a StoryKit that enables young children to physically program, or author, their own StoryRoom experiences [18]. For more information on this work, see http://www.umiacs.umd.edu/ allisond/block/blocks.html.

Acknowledgments

This work has been funded by the European Union's i3 Experimental School Environments initiative, DARPA, and the Institute for Advanced Computer Studies. We would also like to acknowledge current members of our design team: Jack Best, Angela Boltman, Gene Chipman, Cassandra Cosans, Allison Farber, Joe Hammer, Alex Kruskal, Abby Lal, Jade Matthews, Thomas Plaisant–Schwenn, Michele Platner, Jessica Porteous, Emily Rhodes, Lisa Sherman, and Sante Simms.

References

[1] Houman Alborzi, Allison Druin, Jaime Montemayor, Michele Platner, Jessica Porteous, Lisa Sherman, Angela Boltman, Gustav Taxen, Jack Best, Joe Hammer, Alex Kruskal, Abby Lal, Thomas Plaisant-Schwenn, Lauren Sumida, Rebecca Wagner, and James Hendler. Designing storyrooms: Interactive storytelling spaces for children. In *Proceedings of Designing Interactive Systems (DIS-2000)*, pages 95–104. ACM Press, 2000.

[2] Hugh Beyer and Karen Holtzblatt. *Contextual design: defining customer–centered systems*. Morgan Kaufmann, San Francisco, California, 1998.

[3] Aaron Bobick, Stephen S. Intille, James W. Davis, Freedom Baird, Claudio S. Pinhanez, Lee W. Campbell, Yuri A. Ivanov, Arjan Schutte, and Andrew Wilson. The kidsroom: A perceptually-based interactive and immersive story environment. In *PRESENCE: Teleoperators and Virtual Environments*, pages 367–391, August 1999.

[4] R. Peter Bonasso, R. James Firby, Erann Gat, David Kortenkamp, David Miller, and M Slack. Experiences with architecture for intelligent, reactive agents. *Journal of Experimental and Theoretical Artificial Intelligence*, pages 237–256, 1997.

[5] Cynthia Breazeal. A motivational system for regulating human-robot interaction. In *Proceedings of AAAI'98*, pages 126–131. AAAI Press, 1998.

[6] Joseph Bruchac. *Survival this way: Interviews with American Indian poets*. University of Arizona Press, Tuscson, Arizona, 1987.

[7] Allison Druin. Cooperative inquiry: Developing new technologies for children with children. In *Proceedings of Human Factors in Computing Systems (CHI 99)*. ACM Press, 1999.

[8] Allison Druin. The role of children in the design of new technology. Technical Report UMIACS–TR–99–53, UMIACS, 1999.

[9] Allison Druin, Ben Bederson, Juan Pablo Hourcade, Lisa Sherman, Glenda Revelle, Michele Platner, and Stacy Weng. Designing a digital library for young children: An intergenerational partnership. In *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL 2001)*, 2001.

[10] Allison Druin, Jaime Montemayor, James Hendler, Britt McAlister, Angela Boltman, Eric Fiterman, Aurelie Plaisant, Alex Kruskal, Hanne Olsen, Isabella Revett, Thomas Plaisant-Schwenn, Lauren Sumida, and Rebecca Wagner. Designing pets: A personal

electronic teller of stories. In *Proceedings of Human Factors in Computing Systems (CHI'99)*. ACM Press, 1999.

[11] Allison Druin and Ken Perlin. Immersive environments: A physical approach to the computer interface. In *Proceedings of Human Factors in Computing Systems (CHI 94)*, volume 2, pages 325–326. ACM Press, 1994.

[12] Phil Frei, Victor Su, Bakhtiar Mikhak, and Hiroshi Ishii. Curlybot: Designing a new class of computational toys. In *Proceedings of Human Factors in Computing Systems (CHI 2000)*, pages 129–136. ACM Press, 2000.

[13] Masahiro Fujita and Hiroaki Kitano. Development of an autonomous quadruped robot for robot entertainment. *Autonomous Robots*, 5(1):7–18, 1998.

[14] Theodore Geisel. *The Sneetches, and Other Stories*. Random House, New York, 1961.

[15] Robert Franklin Gish. *Beyond bounds: Cross–Cultural essays on Anglo, American Indian, and Chicano literature*. University of New Mexico Press, Albuquerque, NM, 1996.

[16] Fred Martin, Bakhtiar Mikhak, Mitchel Resnick, Brian Silverman, and Robbie Berg. To mindstorms and beyond: Evolution of a construction kit for magical machines. In Allison Druin and James Hendler, editors, *Robots for kids: New technologies for learning*. Morgan Kaufmann, San Francisco CA, 2000.

[17] Fred G. Martin. The handy board technical reference. URL http://el.www.media.mit.edu/projects/handy-board/techdocs/hbmanual.pdf, 1998.

[18] Jaime Montemayor. Physical programming: Software you can touch. In *Proceedings of Human Factors in Computing Systems, Extended Abstracts of Doctoral Consortium (CHI 2001)*. ACM Press, 2001.

[19] Jaime Montemayor, Allison Druin, and James Hendler. Pets: A personal electronic teller of stories. In Allison Druin and James Hendler, editors, *Robots for kids: New technologies for learning*, pages 367–391. Morgan Kaufmann, San Francisco CA, 2000.

[20] Simon J. Ortiz. *Speaking for generations: Native writers on writing*. University of Arizona Press, Tuscson, AR, 1998.

[21] Seymour Papert. *Mindstorms: Children, computers and powerful ideas*. Basic Books, New York, 1980.

[22] Catherine Plaisant, Allison Druin, Cori Lathan, Kapil Dakhane, Kris Edwards, Jack Maxwell Vice, and Jaime Montemayor. A storytelling robot for pediatric rehabilitation. In *Proceedings of ASSETS'2000*. ACM Press, 2000.

[23] Mitchel Resnick, Fred Martin, Robbie Berg, Rick Borovoy, Vanessa Colella, Kwin Kramer, and Brian Silverman. Digital manipulatives: New toys to think with. In *Proceedings of Human Factors in Computing Systems (CHI 98)*, pages 281–287. ACM Press, 1998.

[24] R. J. Semper. Science museums as environments for learning. *Physics Today*, pages 50–56, November 1990.

[25] Erik Strommen. When the interface is a talking dinosaur: Learning across media with actimates barney. In *Proceedings of Human Factors in Computing Systems (CHI 98)*, pages 288–295. ACM Press, 1998.

[26] Marina Umaschi. Soft toys with computer hearts: Building personal storytelling environments. In *Proceedings of Extended Abstracts of Human Factors in Computing Systems (CHI 97)*, pages 20–21. ACM Press, 1997.

Chapter 26

SOCIALLY INTELLIGENT AGENTS IN EDUCATIONAL GAMES

Cristina Conati and Maria Klawe
University of British Columbia

Abstract We describe preliminary research on devising intelligent agents that can improve the educational effectiveness of collaborative, educational computer games. We illustrate how these agents can overcome some of the shortcomings of educational games by explicitly monitoring how students interact with the games, by modeling both the students' cognitive and emotional states, and by generating calibrated interventions to trigger constructive reasoning and reflection when needed.

1. Introduction

Several authors have suggested the potential of video and computer games as educational tools. However empirical studies have shown that, although educational games are usually highly engaging, they often do not trigger the constructive reasoning necessary for learning [4] [12]. For instance, studies performed by the EGEMS (Electronic Games for Education in Math and Science) project at the University of British Columbia have shown that the tested educational games were effective only when coupled with supporting classroom activities, such as related pencil and paper worksheets and discussions with teachers. Without these supporting activities, despite enthusiastic game playing, the learning that these games generated was usually rather limited [12].

An explanation of these findings is that it is often possible to learn how to play an educational game effectively without necessarily reasoning about the target domain knowledge [4]. Insightful learning requires meta-cognitive skills that foster conscious reflection upon one's actions [6], but reflective cognition is hard work. Possibly, for many students the high level of engagement triggered by the game acts as a distraction from reflective cognition, especially when the game is not integrated with external activities that help ground the game experience into the learning one. Also, educational games are usually highly exploratory in nature, and empirical studies on exploratory learning environments [16] have shown that they tend to be effective only for those students that already possess the meta-cognitive skills necessary to learn from autonomous exploration (such as self-monitoring, self-questioning and self-explanation).

In this chapter, we discuss how to improve the effectiveness of educational games by relying on socially intelligent agents (SIAs). These agents are active game characters that can generate tailored interventions to stimulate students' learning and engagement, by taking into account the student's cognitive states (e.g., as knowledge, goals and preferences), as well as the student's meta-cognitive skills (e.g., learning capabilities) and emotional reactions.

2. SIAs as Mediators in Educational Games

We argue that the effectiveness of educational games can be increased by providing them with the capability to (i) explicitly monitor how students interact with and learn from the games; (ii) generate calibrated interventions to trigger constructive reasoning and reflection when needed.

However, this must be done without interfering with the factors that make games fun and enjoyable, such as a feeling of control, curiosity, triggering of both intrinsic and extrinsic fantasies, and challenge [12]. Thus, it is not sufficient to provide educational games with the knowledge that makes more traditional Intelligent Tutoring Systems effective for learning: an explicit representation of the target cognitive skills, of pedagogical knowledge and of the student's cognitive state. It is fundamental that the educational interventions be delivered within the spirit of the game, by characters that (i) are an integral part of the game plot; (ii) are capable of detecting students' lack of engagement, in addition to lack of learning; (iii) know how to effectively intervene to correct these negative emotional and cognitive states.

Basically, these characters must play, in the context of the game, the mediating role that teachers and external instructional activities have played during the most successful evaluations of the EGEMS prototypes. The requirement that these agents be socially intelligent is further enforced by the fact that we are currently interested in investigating the educational potential of multi-player computer games to support collaborative learning. In the last few years there has been increasing research on animated pedagogical agents and there is already empirical evidence of their effectiveness in fostering learning and motivation [17]. Our work extends existing research toward making pedagogical agents more socially apt, by enabling them to take into account users' affective behaviour when adapting their interventions and to engage in effective collaborative interactions.

2.1 SIAs to Support Game-Based Collaborative Learning

Effective collaborative interaction with peers has proven a successful and uniquely powerful learning method [14]. Students learning effectively in groups encourage each other to ask questions, justify their opinions, and reflect upon their knowledge. However, effective group interaction does not just magically happen. It depends upon a number of factors, including the group composition, the task at hand, and the roles that the group members play during the interaction [14]. Some of these factors (such as the composition of the group), need to be taken into account when creating the groups. Others can be enforced during the interaction by a human or artificial agent that oversees the collaboration process and detects when the conditions for effective collaboration are not met. We are working on creating artificial agents that can provide this mediating role within multi-player, multi-activity educational games designed to foster learning through collaboration. As a test-bed for our research we are using Avalanche, one of the EGEMS prototype games, in which four players work together through a set of activities to deal with the problems caused by a series of avalanches in a mountain ski town. Each of the Avalanche activities is designed to foster understanding of a specific set of mathematical and geometrical skills, including number factorisation as well as measurement and estimate of area/volume.

Preliminary pilot studies have shown that the collaborative nature of the game triggers a tremendous level of engagement in the students. However, they also uncovered several problems. First, students seldom read the available on line help and the canned instructions provided within each activity. Thus, students often lose track of the game goals and of the means available to achieve them. Second, often students succeed in the game by learning heuristics that do not necessarily help them learn the target instructional knowledge. Third, the game at times fails to trigger effective collaboration. For instance, students that are not familiar with the other group members tend to be isolated during the interaction, while highly competitive students sometime turn an activity designed to foster collaboration into a competition.

3. A Comprehensive Computational Model of Effective Collaborative Learning

The above examples show that Avalanche can greatly benefit from the addition of SIAs that help students find their way through the game, trigger constructive learning and reflection, and help mediate and structure the collaborative interaction. To succeed in these tasks the agents need to have:
(i) explicit models of the game activities they are associated with, of the emotional states that can influence learning from these activities and of effective collaborative interaction;

(ii) the capability of modeling, from the interaction with the game, the players' cognitive and meta-cognitive skills, along with their emotional states and the status of the collaborative interaction;
(iii) the capability of making intelligent decisions as to when and how to intervene to improve effective collaboration and learning, without compromising the level of motivation and engagement fueled by the game.

3.1 Architecture

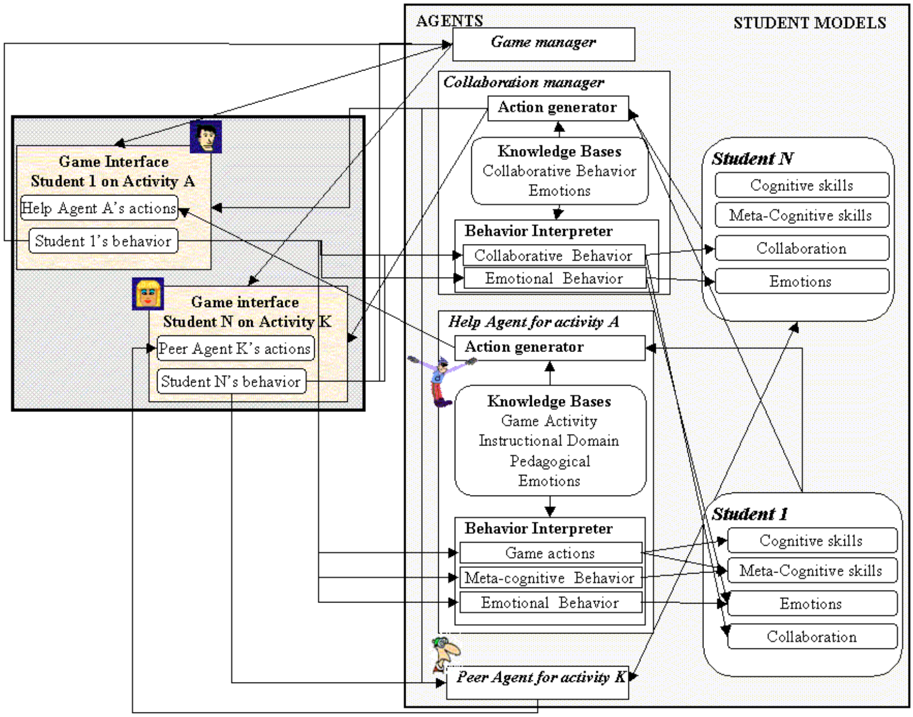


Figure 26.1. Architecture for SIAs in a multi-player, multi-activity educational game

Figure 1 sketches our proposed general architecture underlying the functioning of socially intelligent characters for a multi-player, multi-activity educational game. As students engage in the different activities available within the game, their behavior is monitored by the agents currently involved in the interaction, through their Behavior Interpreters. Each Behavior Interpreter specializes in interpreting actions related to a specific player's behavior (e.g., behavior related to game performance, meta-cognitive skills, collaboration and emotional reaction) and updates the corresponding elements in the student model for that player.

A Game Actions Interpreter, for instance, processes all the student's game actions within a specific activity, to infer information on the student's cognitive and meta-cognitive skills. A Meta-Cognitive Behavior Interpreter tracks all the additional student's actions that can indicate meta-cognitive activity, (e.g., utterances and eye or mouse movements) and passes them to the student model as further evidence on the student's meta-cognitive skills. The agent's action generator then uses the student model and the expertise encoded in the agent's knowledge base (which depend on the agent's pedagogical role) to generate actions that help the student learn better from the current activity.

The agents in the architecture include a Game Manager, the Collaboration Manager and agents related to specific game activities (like Help Agent for activity A and Peer Agent for activity K in Figure 1). The Game Manager knows about the structure of the game and guides the students through its activities. The Collaboration Manager is in charge of orchestrating effective collaborative behavior. As shown in Figure 1, its Behavior Interpreter captures and decodes all those students' actions that can indicate collaboration or lack thereof, along with the related emotional reactions. The actions that pertain to the Collaboration Manager include selecting an adequate collaboration role and partners for a student within a particular activity. The pool of partners from which the Collaboration Manager can select includes both the other players or the artificial agents (e.g., the Peer Agent selected for Student N in activity K in Figure 1), to deal with situations in which no other player can currently be an adequate partner for a student, because of incompatible cognitive or emotional states.

The artificial agents related to each game activity have expertise that allow them to play specific roles within that activity. So, for instance, a Help Agent (like Help Agent for activity A in Figure 1) has expert knowledge on a given activity, on the emotional states that can influence the benefits of providing help and on how to provide this help effectively. Peer agents, on the other hand, will have game and domain knowledge that is incomplete in different ways, so that they can be selected by the Collaboration Manager to play specific collaborative roles in the activity (e.g., that of a more or less skilled learning companion).

3.2 Student Models

The student models in our architecture are based on the probabilistic reasoning framework of Bayesian networks [10] that allows performing reasoning under uncertainty by relying on the sound foundations of probability theory. One of the main objections to the use of Bayesian networks is the difficulty of assigning accurate network parameters (i.e. prior and conditional probabilities). However, even when the parameters cannot be reliably specified by experts or learned from data, providing estimates for them allows the designer to clearly define the assumptions the model must rely upon and to revise the assumptions by trial and error on the model performance. Thus, we believe that Bayesian networks provide an appropriate formalism to model and integrate in a principled way the multiple sources of uncertainty involved in monitoring a student's cognitive and emotional states, and the unfolding of a collaborative interaction.

Modeling cognitive and meta-cognitive skills. Bayesian networks have been extensively used to build user models representing user's knowledge and goals [11]. In [3], we have described how to automatically specify the structure and conditional probabilities of a Bayesian network that models the relations between a user's problem solving behavior and her domain knowledge. In [7], we have extended this work to model learning of instructional material through the meta-cognitive skill known as self-explanation. We plan to adapt this approach to formalize the probabilistic relationships between player's behavior, meta-cognitive skills and learning in the student models for SIAs in educational games.

Modeling collaboration. A preliminary Bayesian model of effective collaborative interaction has been proposed in [13]. The model attempts to trace the progress of group members through different collaborative roles (e.g., leader, observer, critic) by monitoring the actions that they perform on an interface especially designed to reify these roles. We also adopt a role-based approach to model effective collaboration, but we cannot structure and constrain the game interface as in [13], because this kind of highly constrained interaction could compromise the level of fun and engagement that students experience with Avalanche. Hence, we need to devise alternative ways to capture the collaborative roles that students adopt during the interaction. We plan to start by making the adoption of different collaborative roles one of the mandatory game activities, orchestrated by the Collaboration Manager. This will reduce the collaboration-monitoring problem to the problem of verifying that students effectively perform the role they have been assigned. However, as the research proceeds, we hope to also achieve a better understanding of how to monitor and support less constrained collaboration.

Modeling emotions. Since emotional engagement is the element that makes educational games attractive to learners, it is fundamental that this variable be accurately monitored and taken into account by SIAs for these games. Starting from existing research on the structure of emotions [1], we are working on a general Bayesian student model to represent relevant emotional states (such as frustration, boredom and excitement) and their dynamics, as they are influenced by the interaction with an educational game, by the SIAs interventions and by the player's personality [2]. The formalization includes a theory of how the players' emotions can be detected, based on current research on how to measure emotional reactions through bodily expressions such as facial expressions, vocal intonation, galvanic skin response and heart rate [15].

3.3 Action Generators

The action generator for each SIA in the game relies on a decision-theoretic model of decision-making predicting that agents act so as to maximize the expected utility of their actions [9]. Other researchers have started adopting a decision theoretic approach to regulate the behavior of interactive desktop assistants [8] and of an intelligent tutor to support coached problem solving [5].

In our architecture, the function representing an agent's preferences in terms of utility values depends on the role of the agent in the game. So, for instance, the Collaboration Manager will act so as to maximize students' learning as well as their collaborative behavior. A Help Agent will act to maximize the student's understanding of a specific activity, while an agent in charge of eliciting a specific meta-cognitive skill will select actions that maximize this specific outcome. All the agents will also include in their utility functions the goal of maintaining the student's level of fun and engagement above a given threshold, although the threshold may vary with the agent's role. The action generators' decision-theoretic models can be represented as influence diagrams [9], an extension of Bayesian networks devised to model rational decision making under uncertainty. By using influence diagrams, we can compactly specify how each SIA's action influences the relevant elements in the Bayesian student model, such as the player's cognitive and emotional states. We can also encode the agent's utility function in terms of these states, thus providing each agent with a normative theory of how to intervene in the students' game playing to achieve the best trade-off between engagement and learning.

4. Conclusions

We have presented a preliminary architecture to improve the effectiveness of collaborative educational games. The architecture relies on the usage of socially intelligent agents that calibrate their interventions by taking into account not only the students' cognitive states, but also their emotional states and the unfolding of collaborative interactions within the game. We propose to rely on Bayesian networks and influence diagrams to provide our agents with a principled framework for making informed decisions on the most effective interventions under the multiple sources of uncertainty involved in modelling interaction and learning in multi-player, multi-activity educational game.

References

[1] A. Orthony and G.L. Clore and A. Collins. *The cognitive structure of emotions*. Cambridge University Press, Cambridge, England, 1988.

[2] C. Conati. Modeling Users' Emotions to Improve Learning with Educational Games. In *Proceedings of the 2001 AAAI Fall Symposium Intelligent and Emotional Agents II*, pages 31–36. AAAI Press, Menlo Park, U.S.A.

[3] C. Conati and A. Gertner and K. VanLehn and M. Druzdzel. On-line student modeling for coached problem solving using Bayesian networks. In A. Jameson et al., editor, *User Modeling: Proc. Sixth Int. Conf., UM97*, pages 231–242. Springer, New York, 1997.

[4] C. Conati and J. Fain Lehman. EFH-Soar: Modeling education in highly interactive microworlds. In *Lecture Notes in Artificial Intelligence*, pages 47–58. Springer-Verlag, New York, 1993.

[5] C. Murray and K. VanLehn. DT Tutor: A decision-theoretic dynamic approach for optimal selection of tutorial actions. In *Lecture Notes in Computer Science: Intelligent Tutoring Systems, 5th International Conference*, pages 153–162. Springer, 2000.

[6] C. Conati and K. VanLehn. Toward Computer-Based Support of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation. *International Journal of Artificial Intelligence in Education*, 11(4):289–415, 2000.

[7] C. Conati and K. VanLehn. Providing adaptive support to the understanding of instructional material. In *Proceedings of IUI 2001, International Conference on Intelligent User Interfaces, Santa Fe, New Mexico, USA*, pages 41–48. ACM Press, 2001.

[8] E. Horvitz. Principles of mixed initiative interaction. In *CHI '99, ACM SIGCHI Conf. on Human Factors in Computing Systems., Pbgh, Pa*, pages 159–166. ACM Press, 1999.

[9] M. Henrion, J. Breeze, and E. Horvitz. Decision Analysis and Expert Systems. *AI Magazine*, Winter '91:64–91, 1991.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.

[11] A. Jameson. Numerical uncertainty management in user and student modeling: An overview of systems and issues. *User Modeling and User-Adapted Int.*, 5:193–251, 1995.

[12] M. Klawe. When Does The Use Of Computer Games And Other Interactive Multimedia Software Help Students Learn Mathematics? In *NCTM Standards 2000 Technology Conference, Arlington, VA*, 1998.

[13] M. Singley and P. Fairwater. Team Tutoring Systems: Reifying Roles in Problem Solving. In C. Hoadley and J. Roschelle, editor, *CSCL '99, Stanford, California*, pages 538–549. Lawrence Erlbaum Associates, Hillsdale, NJ, 1999.

[14] P. Dillenbourg and M. Baker and A. Blaye and C. O' Malley. The evolution of research on collaborative learning. In E. Spada and P. Reiman, editors, *Learning in Humans and Machine: Towards an interdisciplinary learning science*, pages 189–211. 1996.

[15] R. Picard. *Affective Computing*. M.I.T. Press, Cambridge, Massachusetts, 1997.

[16] V. J. Shute. A comparison of learning environments: All that glitters... In S.P.L. and S.J. Derry, editor, *Computers as Cognitive Tools*, pages 47–73. LEA, Hillsdale, NJ, 1993.

[17] W. L. Johnson and J.W. Rickel and J.C. Lester. Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, 11:47–78, 2000.

Chapter 27

TOWARDS INTEGRATING PLOT AND CHARACTER FOR INTERACTIVE DRAMA

Michael Mateas and Andrew Stern

Computer Science Department, Carnegie Mellon University and www.interactivestory.net

Abstract The authors are currently engaged in a three year collaboration to build an interactive story world integrating believable agents and interactive plot. This paper provides a brief description of the project goals and design requirements, discusses the problem of autonomy in the context of story-based believable agents, and describes an architecture that uses the dramatic beat as a structural principle to integrate plot and character.

1. Introduction

Interactive drama concerns itself with building dramatically interesting virtual worlds inhabited by computer-controlled characters, within which the user (hereafter referred to as the player) experiences a story from a first person perspective [7]). Over the past decade there has been a fair amount of research into believable agents, that is, autonomous characters exhibiting rich personalities, emotions, and social interactions ([12]; [8]; [5]; [4]; [9]; [1]). There has been comparatively little work, however, exploring how the local, reactive behavior of believable agents can be integrated with the more global, deliberative nature of a story plot, so as to build interactive, dramatic worlds ([16]; [2]). The authors are currently engaged in a three year collaboration to build an interactive story world integrating believable agents and interactive plot. This paper provides a brief description of the project goals and design requirements, discusses the problem of autonomy in the context of story-based believable agents, and finally describes an architecture that uses the dramatic beat as a structural principle to integrate plot and character.

2. Design Requirements

Artistically complete. The player should have a complete, artistically whole experience.

Animated characters. The characters will be represented as real-time animated figures that can emote, have personality and can speak.

Interface. The player will experience the world from a first-person 3D perspective. The viewpoint is controlled with the keyboard and mouse.

Dialog. Dialog will be the primary mechanism by which a player interacts with characters and influences how the story unfolds. To achieve dialog, the player types out text that is visible on screen; the computer characters' dialog is spoken speech with simultaneously displayed text. The conversation discourse is real-time; that is, if the player is typing, it is as if they are speaking those words in (pseudo) real-time. The system should be very robust when responding to inappropriate and unintelligible input.

Interactivity and plot. The player's actions should have a significant influence on what events occur in the plot, which are left out, and how the story ends. The plot should be generative enough that it supports replayability. Only after playing the experience 6 or 7 times should the player begin to feel they have "exhausted" the interactive story. In fact, full appreciation of the experience requires the story be played multiple times.

Short one-act play. We want to design an experience that provides the player with 15 minutes of emotionally intense, tightly unified, dramatic action.

Relationships. The story should be about the emotional entanglements of human relationships. Our story is a domestic drama in which the relationship of a married couple, Grace and Trip, falls apart during an innocent evening visit by the Player.

Three characters. The story should have three characters, two controlled by the computer and one controlled by the player.

The player should not be over-constrained by a role. The amount of non-interactive exposition describing the player's role should be minimal.

Distributable. The system will be implemented on a platform that is reasonably distributable, with the intention of getting the interactive experience into the hands of as many people as possible.

For more details, see [13].

3. Autonomy and Story-Based Believable Agents

Most work in believable agents has been organized around the metaphor of strong autonomy. Such an agent chooses its next action based on local perception of its environment plus internal state corresponding to the goals and possibly the emotional state of the agent. Using autonomy as a metaphor driving the design of believable agents works well for believable agent applications in which a single agent is facilitating a task, such as instructing a student ([9]), or giving a presentation ([6]), or in entertainment applications in which a user develops a long-term relationship with the characters by "hanging-out" with them ([1]). But for believable agents used as characters in a story world, strong autonomy becomes problematic. Knowing which action to take at any given time depends not just on the private internal state of the agent plus current world state, but also on the current story state, including the entire past history of interactions building on each other towards some end. The global nature of story state is inconsistent with the notion of an autonomous character that makes decisions based only on private goal and emotion state and local sensing of the environment.

Only a small amount of work has been done on the integration of story and character. This work has preserved the strong autonomy of the characters by architecturally dividing the responsibility for state maintenance between a drama manager, which is responsible for maintaining story state, and the believable agents, which are responsible for maintaining character state and making the moment-by-moment behavior decisions ([16]; [2]). These two components communicate via a narrow-bandwidth, one-directional interface flowing from drama manager to agent. The messages sent across this interface consist of goals that characters should assume or perhaps specific actions they should perform. The character is still responsible for most of the decision making.

This architecture makes several assumptions regarding the nature of interactive drama and believable agents: drama manager decisions are infrequent, the internal structure of the believable agents can be reasonably decoupled from their interaction with the drama manager, and multiple-character coordination is handled within the agents. Let's explore each of these assumptions.

Infrequent guidance of strongly autonomous believable agents means that most of the time, behavior selection for the believable agents will occur locally, without reference to any (global) story state. The drama manager will intervene to move the story forward at specific points; the rest of the time the story will be "drifting," that is, action will be occurring without explicit attention to story movement. Weyhrauch ([16]) does state that his drama manager was designed for managing the sequencing of plot points, that is, for guiding characters so as to initiate the appropriate next scene necessary to make the next plot point happen (whatever plot point has been decided by the drama manager). Within a scene, some other architectural component, a "scene manager," would be necessary to manage the playing out of the individual scene. And this is where the assumption of infrequent, low-bandwidth guidance becomes violated. As is described in the next section, the smallest unit of story structure within a scene is the beat, a single action/reaction pair. The scene-level drama manager will thus need to continuously guide the autonomous decision making of the agent. This frequent guidance from the drama manager will be complicated by the fact that low-bandwidth guidance (such as giving a believable agent a new goal) will interact strongly with the moment-by-moment internal state of the agent, such as the set of currently active goals and behaviors, leading to surprising, and usually unwanted, behavior. In order to reliably guide an agent, the scene-level drama manager will have to engage in higher-bandwidth guidance involving the active manipulation of internal agent state (e.g. editing the currently active goal tree). Authoring strongly autonomous characters for story-worlds is not only extra, unneeded work (given that scene-level guidance will need to intervene frequently), but actively makes guidance more difficult, in that the drama manager will have to compensate for the internal decision-making processes (and associated state) of the agent.

As the drama manager provides guidance, it will often be the case that the manager will need to carefully coordinate multiple characters so as to make the next story event happen. For example, it may be important for two characters to argue in such a way as to conspire towards the revelation of specific information at a certain moment in the story. To achieve this with autonomous agents, one could try to back away from the stance of strong autonomy and provide special goals and behaviors within the individual agents that the drama manager can activate to create coordinated behavior. But even if the character author provides these special coordination hooks, coordination is still being handled at the individual goal and behavior level, in an ad-hoc way. What one really wants is a way to directly express coordinated character action at a level above the individual characters.

At this point the assumptions made by an interactive drama architecture consisting of a drama manager guiding strongly autonomous agents have been found problematic. The next section presents a sketch of a plot and character architecture that addresses these problems.

4. Integrating Plot and Character with the Dramatic Beat

In dramatic writing, stories are thought of as consisting of events that turn (change) values ([14]). A value is a property of an individual or relationship, such as trust, love, hope (or hopelessness), etc. A story event is precisely any activity that turns a value. If there is activity – characters running around, lots of witty dialog, buildings and bridges exploding, and so on – but this activity is not turning a value, then there is no story event, no dramatic action. Thus one of the primary goals of an interactive drama system should be to make sure that all activity turns values. Of course these values should be changed in such a way as to make some plot arc happen that enacts the story premise, such as in our case, "To be happy you must be true to yourself".

Major value changes occur in each scene. Each scene is a large-scale story event, such as "Grace confesses her fears to the player". Scenes are composed of beats, the smallest unit of value change. Roughly, a beat consists of one or more action/reaction pairs between characters. Generally speaking, in the interest of maintaining economy and intensity, a beat should not last longer than a few actions or lines of dialog.

4.1 Scenes and Beats as Architectural Entities

Given that the drama manager's primary goal is to make sure that activity in the story world is dramatic action, and thus turns values, it makes sense to have the drama manager use scenes and beats as architectural entities.

In computational terms, a scene consists of preconditions, a description of the value(s) intended to be changed by the scene (e.g. love between Grace and the player moves from low to high), a (potentially large) collection of beats with which to construct the scene, and a description of the arc that the value(s) changed by the scene should follow within the scene. To decide which scene to attempt to make happen next, the drama manager examines the list of unused scenes and chooses the one that has a satisfied precondition and whose value change best matches the shape of the global plot arc.

Once a scene has been selected, the drama manager tries to make the scene play out by selecting beats that change values appropriately. A beat consists of preconditions, a description of the values changed by the beat, success and failure conditions, and a joint plan to coordinate the characters in order to carry out the specific beat.

4.2 The Function of Beats

Beats serve several functions within the architecture. First, beats are the smallest unit of dramatic value change. They are the fundamental building blocks of the interactive story. Second, beats are the fundamental unit of character guidance. The beat defines the granularity of plot/character interaction. Finally, the beat is the fundamental unit of player interaction. The beat is the smallest granularity at which the player can engage in meaningful (having meaning for the story) interaction.

4.3 Polymorphic Beats

The player's activity within a beat will often determine exactly which values are changed by a beat and by how much. For example, imagine that Trip becomes uncomfortable with the current conversation - perhaps at this moment in the story Grace is beginning to reveal problems in their relationship – and he tries to change the topic, perhaps by offering to get the player another drink. The combination of Grace's line of dialog (revealing a problem in their relationship), Trip's line of dialog (attempting to change the topic), and the player's response is a beat. Now if the player responds by accepting Trip's offer for a drink, the attempt to change the topic was successful, Trip may now feel a closer bond to the player, Grace may feel frustrated and angry with both Trip and the player, and the degree to which relationship problems have been revealed does not increase. On the other hand, if the player directly responds to Grace's line, either ignoring Trip, or perhaps chastising Trip for trivializing what Grace said, then the attempt to change the topic was unsuccessful, Trip's affiliation with the player may decrease and Grace's increase, and the degree to which relationship problems have been revealed increases. Before the player reacts to Grace and Trip, the drama manager does not know which beat will actually occur. While this polymorphic beat is executing, it is labelled "open." Once the player "closes" the beat by responding, the drama manager can now update the story history (a specific beat has now occurred) and the rest of the story state (dramatic values, etc.).

4.4 Joint Plans

Associated with each beat is a joint plan that guides the character behavior during that beat. Instead of directly initiating an existing goal or behavior within the character, the drama manager hands the characters new plans (behaviors) to be carried out during this beat. These joint plans describe the coordinated activity required of all the characters in order to carry out the beat. Multi-agent coordination frameworks such as joint intentions theory ([15]) or shared plans ([3] provide a systematic analysis of all the synchronization issues that arise when agents jointly carry out plans. Tambe ([17]) has built an agent architecture providing direct support for joint plans. His architecture uses the more formal analyses of joint intentions and shared plans theory to provide the communication requirements for maintaining coordination. We propose modifying the reactive planning language Hap ([11]; [10]), a language specifically designed for the authoring of believable agents, to include this coordination framework.

Beats will hand the characters joint plans to carry out which have been designed to accomplish the beat. This means that most (perhaps all) of the high level goals and plans that drive a character will no longer be located within the character at all, but rather will be parcelled out among the beats. Given that the purpose of character activity within a story world is to create dramatic action, this is an appropriate way of distributing the characters' behavior. The character behavior is now organized around the dramatic functions that the behavior serves, rather than organized around a conception of the character as independent of the dramatic action. Since the joint plans associated with beats are still reactive plans, there is no loss of character reactivity to a rapidly changing environment. Low-level goals and behaviors (e.g. locomotion, ways to express emotion, personality moves, etc.) will still be contained within individual characters, providing a library of character- specific actions available to the higher-level behaviors handed down by the beats.

5. Conclusion

In this paper we described the project goals of a new interactive drama project being undertaken by the authors. A major goal of this project is to integrate character and story into a complete dramatic world. We then explored the assumptions underlying architectures which propose that story worlds should consist of strongly autonomous believable agents guided by a drama manager, and found those assumptions problematic. Finally, we gave a brief sketch of our interactive drama architecture, which operationalizes structures found in the theory of dramatic writing, particularly the notion of organizing dramatic value change around the scene and the beat.

References

[1] A. Stern and A. Frank and B. Resner. Virtual Petz: A hybrid approach to creating autonomous, lifelike Dogz and Catz. In *Proceedings of the Second International Conference on Autonomous Agents*, pages 334–335. AAAI Press, Menlo Park, California, 1998.

[2] B. Blumberg and T. Galyean. Multi-level Direction of Autonomous Creatures for Real-Time Virtual Environments. In *Proceedings of SIGGRAPH 95*, 1995.

[3] B. Grosz and S. Kraus. Collaborative plans for complex group actions. *Artificial Intelligence*, 86:269–358, 1996.

[4] B. Hayes-Roth and R. van Gent and D. Huber. Acting in character. In R. Trappl and P. Petta, editor, *Creating Personalities for Synthetic Actors*. Springer-Verlag, Berlin, New York, 1997.

[5] B. Blumberg. *Old Tricks, New Dogs: Ethology and Interactive Creatures*. PhD thesis, MIT Media Lab, 1996.

[6] E. Andre and T. Rist and J. Mueller. Integrating Reactive and Scripted Behaviors in a Life-Like Presentation Agent. In *Proceedings of the Second International Conference on Autonomous Agents (Agents '98)*, pages 261–268, 1998.

[7] J. Bates. Virtual Reality, Art, and Entertainment. *Presence: The Journal of Teleoperators and Virtual Environments*, 1:133–138, 1992.

[8] J. Bates and A.B. Loyall and W. S. Reilly. Integrating Reactivity, Goals, and Emotion in a Broad Agent. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, Bloomington, Indiana, July*, 1992.

[9] J. Lester and B. Stone. Increasing Believability in Animated Pedagogical Agents. In *Proceedings of the First International Conference on Autonomous Agents, Marina del Rey, California*, pages 16–21, 1997.

[10] A. B. Loyall. *Believable Agents*. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1997. CMU-CS-97-123.

[11] A. B. Loyall and J. Bates. Hap: A Reactive, Adaptive Architecture for Agents. Technical Report CMU-CS-91-147, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1991.

[12] M. Mateas. An Oz-Centric Review of Interactive Drama and Believable Agents. In M. Wooldridge and M. Veloso, editor, *AI Today: Recent Trends and Developments. Lecture Notes in AI Number 1600*. Springer-Verlag, Berlin, New York, 1999.

[13] M. Mateas and A. Stern. Towards Integrating Plot and Character for Interactive Drama. In *Working notes of the Socially Intelligent Agents: Human in the Loop Symposium, 2000 AAAI Fall Symposium Series*. AAAI Press, Menlo Park, California, 2000.

[14] R. McKee. *Story: Substance, Structure, Style, and the Principles of Screenwriting*. Harper Collins, New York, 1997.

[15] P. Cohen and H. Levesque. Teamwork. *Nous*, 35, 1991.

[16] P. Weyhrauch. *Guiding Interactive Drama*. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1997. Tech report CMU-CS-97-109.

[17] M. Tambe. Towards Flexible Teamwork. *Journal of Artificial Intelligence Research*, 7:83–124, 1997.

Chapter 28

THE COOPERATIVE CONTRACT IN INTERACTIVE ENTERTAINMENT

R. Michael Young

Liquid Narrative Group, North Carolina State University

Abstract Interactions with computer games demonstrate many of the same social and communicative conventions that are seen in conversations between people. I propose that a co-operative contract exists between computer game players and game systems (or their designers) that licenses both the game players' and the game designers' understanding of what components of the game mean.

As computer and console games become more story-oriented and interactivity within these games becomes more sophisticated, this co-operative contract will become even more central to the enjoyment of a game experience. This chapter describes the nature of the co-operative contract and one way that we are designing game systems to leverage the contract to create more compelling experiences.

1. Introduction

When people speak with one another, they co-operate. Even when we argue, we are collaborating together to exchange meaning. In fact, we agree on a wide range of communicative conventions; without these conventions, it would be impossible to understand what each of us means when we say something. This is because much of what we mean to communicate is conveyed not by the explicit propositional content of our utterances, but by the implicit, intentional way that we rely or fail to rely upon conventions of language use when we compose our communication.

Across many media, genres and communicative contexts, the expectation of co-operation acts much like a contract between the participants in a communicative endeavor. By establishing mutual expectations about how we'll be using the medium of our conversation, the contract allows us to eliminate much of the overhead that communication otherwise would require. Our claim is that this compact between communicative participants binds us just as strongly when we interact with computer games as when we interact with each other in more conventional conversational settings. Further, by building systems that are sensitive to the nature of this co-operative contract, it's the goal of our research to enable the creation of interactive narratives that are more engaging as well as more compelling than current state-of-the-art interactive entertainment.

2. Cooperative Discourse Across Genre and Across Media

H. P. Grice, the philosopher of language, characterized conversation as a co-operative process [3] and described a number of general rules, called the *Maxims of Conversation*, that a co-operative speaker follows. According to Grice, speakers select what they say in obedience to these rules, and hearers draw inferences about the speaker's meaning based on the assumption that these rules guide speakers' communication. Grice's Co-operative Principle states:
"Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."
From this very general principle follow four maxims of conversation:

- **The Maxim of Quantity: Make your contribution as informative as required but no more so.**
- **The Maxim of Quality: Try to make your contribution one that is true.**
- **The Maxim of Relation: Be relevant.**
- **The Maxim of Manner: Be perspicuous.**

The Co-operative Principle and its maxims license a wide range of inferences in conversation that are not explicitly warranted by the things that we say. Consider the following exchange:

Bob: How many kids do you have?
Frank: I've got two boys.

In this exchange, Bob relies upon the Maxim of Quantity to infer that Frank has only two children, even though Frank did not say that he had two and only two boys and, furthermore, no girls. For Frank to respond as he does should he have two boys and two girls at home would be uncooperative in a Gricean sense precisely because it violates our notions of what can be inferred from what is left unsaid.

This is just one example of how meaning can be conveyed without being explicitly stated, simply based on an assumption of co-operativity. This reliance upon co-operation is also observable in contexts other than person-to-person communication. For instance, the comprehension of narrative prose fiction relies heavily on inferences made by a reader about the author's intent. Consider the following passage, suggested by the experiments in [9]. James Bond has been captured by criminal genius Blofeld and taken at gunpoint to his hideout.

> James' hands were quickly tied behind his back, but not before he deftly slid a rather plain-looking black plastic men's comb into the back pocket of his jump suit. Blofeld's man gave him a shove down the hallway towards the source of the ominous noises that he'd heard earlier.

In the passage above, the author makes an explicit reference to the comb in James' pocket. As readers, we assume that this information will be central to some future plot element (e.g., the comb will turn out to be a laser or a lock pick or a cell phone) - why else would the author have included it? So we set to work at once anticipating the many ways that James might use the "comb" to escape from what seems a serious predicament. When the comb later turns out to be as central as we suspected, we're pleased that we figured it out, but the inference that we made was licensed only by our assumption that the author was adhering to the Maxim of Relevance. In fact, Relevance comes to play so often in narrative that its intentional violation by an author has a name of its own: the red herring.

This type of co-operative agreement exists in other, less conventional communicative contexts as well. Film, for instance, also relies on the same communicative principles [2]. As one example, when the location of action in a film changes from Place A to Place B, filmmakers often insert an external shot of Place B after the action at Place A ends. Called an *establishing shot*, this inserted footage acts as a marker for the viewer, helping her to understand the re-location of the action without breaking the narrative flow by making the transition explicit.

3. A Cooperative Contract for Interactive Stories

For the designer of a narrative-oriented game that allows substantive user interaction, the greatest design challenge revolves around the maintenance of the co-operative contract, achieved by the effective distribution of control between the system and its users. If a game design removes all control from the user, the resulting system is reduced to conventional narrative forms such as literature or film. As we've discussed above, well-established conventions in these media provide clear signals to their audience, but provide for no interaction with the story. Alternatively, if a game design provides the user with complete control, the narrative coherence of a user's interaction is limited by her own knowledge and abilities, increasing the likelihood that the user's own actions in the game world will, despite her best efforts, fail to mesh with the storyline.

Most interactive games have taken a middle ground, specifying at design-time sets of actions from which the user can choose at a fixed set of points through a game's story. The resulting collection of narrative paths is structured so that each path provides the user with an interesting narrative experience and ensures that the user's expectations regarding narrative content are met. This approach, of course, limits the number and type of stories that can be told inside a single game.

In our work on interactive narrative in the Liquid Narrative research group at North Carolina State University, our approach is to provide a mechanism by which the narrative structure of a game is generated at execution time rather than at design time, customized to user preferences and other contextual factors. The programs that we use to create storylines build models of the story plots that contain a rich causal structure – all causal relationships between actions in the story are specifically marked by special annotations. We put the annotations to good use during gameplay every time that a user attempts to perform an action. As a user attempts to change the state of the world (e.g., by opening a door, picking up or dropping an artifact), a detailed internal model of that action is checked against the causal annotations present in the story. As I describe in more detail below, if the successful completion of the user's action poses a threat to any of the story structure, the system responds to ensure that the actions of the user are integrated as best as possible into the story context.

It is the interactive nature of a computer game that contributes most strongly to the unique sense of agency that gamers experience in the narratives that the game environment supports. But the role of the gamer in a typical computer game is not one of director, but rather of lead character. She does not enter the game world omniscient and omnipotent, but experiences the story that unfolds around her character simultaneously through the eyes of an audience member, the eyes of a performer and through the eyes of her character itself. To uphold her portion of the co-operative contract, she must act well her part, given her limited perceptions and capability to change the game environment.

Consequently, the system creating the storyline behind the scenes must bear most of the responsibility for maintaining the work product of the collaboration, i.e., a coherent narrative experience. To do this, it must plan out ahead of time an interesting path through the space of plot lines that might unfold within the game's storyworld. In addition, the game itself must keep constant watch over the story currently unfolding, lest the user, either by ignorance, accident or maliciousness, deviate from the charted course.

Fortunately, all aspects of a user's activity with the game system, from the graphical rendering of the world to the execution of the simplest of user actions, are controlled (well at least, they're controllable). It is the mediated nature of the interaction between player and game environment that provides us with the hook needed to make the game system co-operative in a Gricean sense. That is, to provide the user with a sense of agency while still directing the flow of a story around the user's (possibly unpredicted) actions.

To support this mediation we are developing a system that sits behind the scenes of a computer game engine, directing the unfolding action while monitoring and reacting to all user activity. The system, called *Mimesis*[6], uses the following components:

1. A declarative representation for action within the environment. This may appear in the type of annotations to virtual worlds suggested by Doyle and Hayes-Roth [4], specifically targeted at the representational level required to piece together plot using plan-based techniques described below.

2. A program that can use this representation to create, modify and maintain a narrative plan, a description of a narrative-structured action sequence that defines all the activity within the game. The narrative plan represents the activities of users, system-controlled agents and the environment itself. This program consists of two parts: an AI planning algorithm such as Longbow [7] and an execution-management component. The planning algorithm constructs plans for user and system interaction that contain such interesting and compelling narrative structure as rising action, balanced conflict between protagonist and antagonist, suspense and foreshadowing. The execution manager issues directives for action to the system's own resources (e.g., the story's system-controlled characters), detects user activities that deviate from the planned narrative and makes real-time decisions about the appropriate system response to such deviations. The response might take the form of re-planning the narrative by modifying the as-yet-unexperienced portions of the narrative plan, or it might take the form of system intervention in the virtual world by preventing the user's deviation from the current plan structure.

3. A theory capable of characterizing plans based on their narrative aspects. This theory informs the program, guiding the construction of plans whose local and global structure are mapped into the narrative structures of conflict, suspense, etc.

4. Conclusions

People interact with systems such as computer games by using many of the same social and communicative conventions that are seen in interactions between people [8]. I propose that expectations about collaboration between computer game players and game systems (or their designers) that licenses both the game players' and the game designers' understanding of what components of the game *mean*. Consequently, the co-operative nature of the gaming experience sets expectations for the behavior of both the game and its players. As computer and console games become more story-oriented and interactivity within these games becomes more sophisticated, this co-operative contract between game and user will become even more central to the enjoyment of a game experience.

The basic building blocks of story and plot — autonomous characters, actions and their causal relationships — are not new to researchers in Artificial Intelligence. These notions are the stuff that makes up most representational schemes in research that deals with reasoning about the physical world. Much of this work has been adapted in the Mimesis architecture to represent the hierarchical and causal nature of narratives identified by narrative theorists [1]. The idea that Grice's Co-operative Principle might be put to use to characterize interactions between people and computers is also not new [5]. But the question of balance between narrative coherence and user control remains an open one, and will not likely be answered by research into human-computer interaction or by modification of conventions carried from over previous entertainment media. It seems more likely that the balance between interactivity and immersion will be established by the concurrent evolution (or by the co-evolution) of the technology of storytelling and social expectations held by the systems' users.

References

[1] Mieke Bal. *Introduction to the Theory of Narrative*. University of Toronto Press, Toronto, Ontario, 1997.

[2] Edward Branigan. *Narrative Comprehension and Film*. Routledge, London and New York, 1992.

[3] H. Paul Grice. Logic and Conversation. In P. Cole and J. L. Morgan, editor, *Syntax and Semantics, vol. 9, Pragmatics*, pages 113–128. Academic Press, New York, 1975.

[4] Patrick Doyle and Barbara Hayes-Roth. Agents in Annotated Worlds. In *Proceedings of the Second International Conference on Autonomous Agents*, pages 35–40, 1998.

[5] R. Michael Young. Using Grice's Maxim of Quantity to Select the Content of Plan Descriptions. *Artificial Intelligence*, 115:215–256, 1999.

[6] R. Michael Young. An Overview of the Mimesis Architecture: Integrating Intelligent Narrative Control into an Existing Gaming Environment. In *The Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, Stanford, California*, pages 77–81, 2001.

[7] R. Michael Young and Martha Pollack and Johanna Moore. Decomposition and Causality in Partial Order Planning. In *Proceedings of the Second International Conference on Artificial Intelligence and Planning Systems*, pages 188–193, 1994.

[8] Clifford Reeves and Byron Nass. *The Media Equation.* Cambridge University Press, Cambridge, England, 1996.

[9] Richard Gerrig. *Experiencing Narrative Worlds.* Yale University Press, New Haven, Connecticut, 1993.

Chapter 29

PERCEPTIONS OF SELF IN ART AND INTELLIGENT AGENTS

Nell Tenhaaf
Department of Visual Arts, York University, Toronto

Abstract The article discusses the term "embodiment" according to the different meanings it has in contemporary cultural discourse on the one hand, and in Artificial Intelligence or Artificial Life modeling on the other. The discussion serves as a backdrop for analysis of an interactive artwork by Vancouver artist Liz Van der Zaag, "Talk Nice", which behaves like an Intelligent Agent that interacts socially with humans. "Talk Nice" has features corresponding to both conceptions of embodiment, and it elicits further ideas about the significance of those notions for definitions of selfhood.

"Embodiment" has come to mean different things in the realms of cultural discourse about art objects on the one hand, and the development of computational artifacts within Artificial Intelligence (AI) or Artificial Life (Alife) research on the other. In the cultural domain, embodiment tends to refer to either mending or transcending the Cartesian mind-body split that has dominated Western thought since the Enlightenment. In Alife and AI however, it means computationally building agents in such a way that they are responsive to their environment, exhibit complex behaviours, and are autonomous to some degree. For convenience I will here collectively refer to the production of these latter artifacts as research on Intelligent Agents, or IA.

Given the dominance of sight in the history of art and its links with a denigration of the body, embodiment in art and culture most often signifies a reintegration into the aesthetic experience of senses other than the visual. These artistically less familiar senses – for example touch or smell – have come to be thought of as more body-based senses since they require somatic involvement that extends beyond the "disembodied eye". Art objects can be made in such a way as to generate embodiment by appealing to these senses, for example Toronto artist Bill Burns' everyday objects formed from chocolate, made in the 1980s and many of them still extant if not any longer as odorous. Ottawa

(Canada) based artist Catherine Richards appeals to the kinesthetic sense in her installation *Virtual Body* (1993), by having the viewer insert a hand into what appears to be an old-fashioned "magic lantern" type of box. Peering in through a lens on the top of the box, the viewer sees behind their own hand a rapidly moving video pattern on a horizontal screen, which translates into a sense of one's feet moving out from underneath as one's hand seems virtually to fly forward. These kinds of works make a very direct appeal to a fuller human sensorium than traditional works of art.

Complicating this portrait of recent shifts in creativity, virtuality or simulation in art objects is often seen as inspiring disembodiment or even outright obsolescence of the body. The Australian performance artist Stelarc describes his work with body prostheses, penetration of the body with robotic objects, and digitally-controlled muscle stimulation as an obsolescence of the body, although he qualifies this as the Cartesian body that has been thought of as distinct from and controlled by the mind. Some artists and cultural critics argue the opposite, that there is always a sensory experience even in virtual space. Jennifer Fisher describes Montreal artist Char Davies' virtual reality installation *Ephémère* (1998) as notable for "its implications for a haptic aesthetics – the sensational and relational aspects of touch, weight, balance, gesture and movement" [2, pp. 53-54]. This work that requires a headset for the viewer also uses pressure sensors in a vest that respond to the expansion and contraction of respiration (as you inhale, you ascend in the simulated world; as you exhale, you sink), and another set of sensors that move the world in response to the tilt of the spinal axis. It is described as a fully immersive experience, meaning whole-body involvement. However, other writers such as robotics artist Simon Penny propose that the computer itself, with its central role in generating virtuality, reinstates Cartesian duality by disengaging and de-emphasizing the physical body from its simulated brain processes [6, pp. 30-38]. There is no singular way of approaching embodiment in art discourse, but one operative principle is that multi-sensory tends to equal greater embodiment and that this offers a fuller, richer aesthetic experience.

Traditionally, experiencing an art work means that the viewer should ideally understand its impact as a gain in self-awareness. Art is about human nature, its innate features and how it unfolds in the world. In the European tradition art has always been directed toward a profound identification between humanism and selfhood in its impact on a viewer, reinforcing these qualities at the centre of a very large opus of philosophical speculation about the meaning of creativity. This idealized picture of aesthetic response is of course a simplification, since critical understanding of the exchange between viewer and art object in practice has many variations. Much discourse of the past three decades especially has shifted art into the arena of social and political meaning. But it remains individual subjectivity that is most often solicited in both the creation and dissemination of an art experience. The current pursuit of re-embodiment as a creative act answers to the cultural dominance of simulation in image-making, and also makes a direct appeal to emotion – think of the associative power of smell. While emotion has always had a place at the core of aesthetic theory, its manifestation adapts to changing cultural conditions. Greater embodiment in the art experience still implies for the viewer an expansion of the sense of self, through these various solicitations of an integrated somatic, perceptual and intellectual response.

In IA research, embodiment is a quite extensive concept that underlies many of the more "lifelike" features of intelligent agents. The autonomous robotic or computer-based agents of IA research are built to be aware of and interact with their environment, as well as interacting with each other and with humans. The types of possible interactions are broadly defined enough to encompass simple behaviours, usually hard-wired to be adaptive to the immediate environment, as well as complex routines such as learning, some level of intentionality, and other features of emergent, evolved behaviour.

Even if "self" in everyday speech signifies the human ability to attach both intellectual and emotional meaning to a lifetime of accumulated memory, the language that describes characteristics of emergent order such as self-organizing or self-regulating, when applied not to physical processes but to these embodied artificial entities, implies at least in principle a generating of "selfhood." This follows especially from the Alife logic that programmed functions of agents parallel life processes, so that emergent and fully autonomous behaviour would equal alive – which would then entail a sense of self [1]. Equally, there are descriptions from the cultural domain of such a non-anthropocentric idea of self. French theoretician Georges Bataille says, "Even an inert particle, lower down the scale than the animalcula, seems to have this existence for-itself, though I prefer the words inside or inner experience" [3, p. 99]. He does go on to say, though, that this elementary feeling of self is not consciousness of self, that is distinctly human. Thus the two meanings of embodiment, and therefore the kind of experience that a person might have in relation to either an artwork or an IA, at first seem to meet in their privileging of some kind of selfhood. The features of an IA that are an effect of its artificial, non-human self-recognition may very well mirror and enhance the sense of self in a person interacting with it. This would be most ensured by well-developed characteristics of social and emotional intelligence built into the agent, so that interactions with it seem natural.

But while concepts of embodiment are representational issues dependent on the intrinsic qualities of artifacts and how those are conveyed, the investigation of selfhood vis-^-vis these artifacts of research or art practice is necessarily interactive. It is bound up in our relations with them. Given the strong humanist tradition of art and the implicit technological nature of IAs, our relational experience of them ultimately diverges as much as the two notions of embodiment in them differ. Experiencing simulated self-recognition in an IA is likely to not reinforce the sense of self in the human interactor at all, but rather counter it and provoke a relinquishing of selfhood in parallel with the process of recognizing an artificial self. This is because the simulation itself, the technological construction of the IA, situates it within the "ethos" of technology that imposes a possibly dehumanizing but always rationally utilitarian value onto its artifacts [4, pp. 38,50]. Which is to say that ordinary people, more or less unwittingly, experience autonomous artifacts through a disposition of what they wish technology to do for them. They unconsciously attribute to the artifact, as to all technological apparati, the power to satisfy their desires.

An IA will thus have a radically different impact than traditional kinds of art, although it may come closer to paralleling more recent experimental art that pursues re-embodiment by engaging senses other than the visual. Vancouver-based artist Elizabeth Van der Zaag's interactive work *Talk Nice* could be approached and analyzed as the latter, since the viewer is required to sit in a chair and talk through a microphone to a video projection, which then responds to the input. One could argue that the viewer is more physically aware of their own presence in the work because of these features. But *Talk Nice* is more accurately described as an artwork that behaves like an IA. From an IA research point of view, Van der Zaag's speaking/listening system is itself an embodied agent through its ability to interact with humans, so as to calculate and then communicate an assessment of human performance. Once the viewer has crossed the threshold of reluctance (in my case) to speak aloud to a virtual other in a public space, the contest for mastery of the situation – human or machine – begins.

Talk Nice uses SAY (Speak and Yell) software, created by the artist herself, which detects loudness and the pitch at the end of a sentence in the participant's voice. The chair and microphone for participant input are located about ten feet from a video projection that shows two young women seated at a table, plus a floating red ball and a blue bar to the right of this scenario that reflects the pitch change in the participant's voice (Fig. 1), and a red line along the bottom that shows the amplitude or loudness of the voice. Sitting in the chair turns on the microphone, whereupon the girls remark that someone is there and prompt the participant to speak. Their first response, which launches the "coaching sessions," is that the loudness of your voice is okay or not right. But the change in pitch at the very last second of your sentence is what counts, and so the coaching videos continue with help in learning how to speak with an "upism." The interaction is set up as a game: the *Talk Nice* flow chart (Fig. 2) tracks the pathways through learning and subsequent moves into the chat of the Bubble Tea Room and the goal of going to the cool Party.

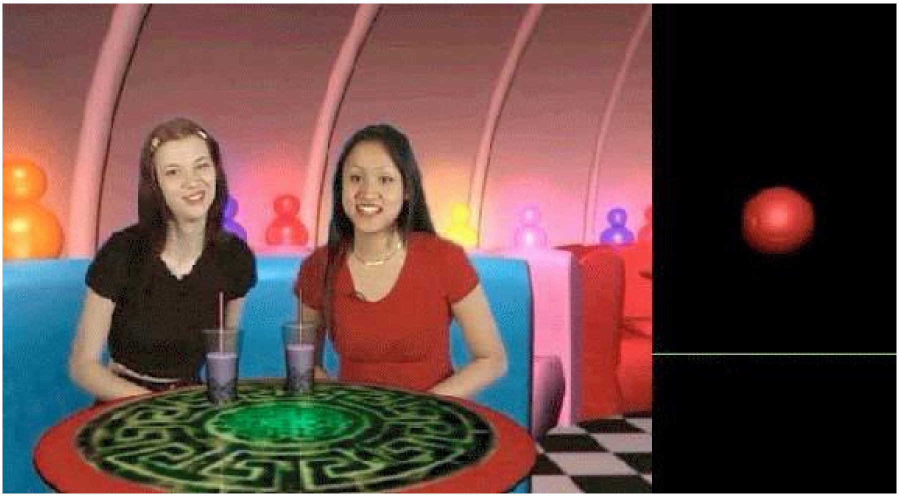


Figure 29.1. Talk Nice display

Talk Nice exhibits social understanding by eliciting and responding to a self-consciousness in the viewer about their speech, their bodily dynamics, and their own mechanisms of understanding. But Van der Zaag says that she is not interested in virtuality (and therefore, one could assume, the autonomy of her agent), or how human relations have been changed by it. Rather, she describes her work as directed toward the changing nature of emotionality in language and strategies for eliciting audience attention to such issues. The technological setup is just a facilitator for an investigation of evolving language exchange among people. Yet this begs the question as to why she would use an artificial, interactive setup to focus on language. It builds into the work an implication that mimicry through the pervasiveness of electronic media plays an important part in transformations of language, specifically the "upism" that the participant is to learn. Although the key practice phrase for learning how to speak this way is the now broadly familiar, "I'm a Canadian, eh?" with its upward lilt on that last word, my sense is that the popular use of this mode of speech spread via TV from the Valley Girls of California in the eighties. Van der Zaag naturalizes these kinds of subtle changes in usage, by setting up her software agent as an extension of human exchange rather than foregrounding ideas about autonomy or emergence. After all, SAY only hears *how* you say it, not what you say.

But more to the point, whatever the entanglement here between the participant, the agent, and the social history of language, and whether we consider the *Talk Nice* system from an IA or artwork point of view, the agent nonetheless has a lot of authority. It is perhaps even more authoritarian than if it tried to

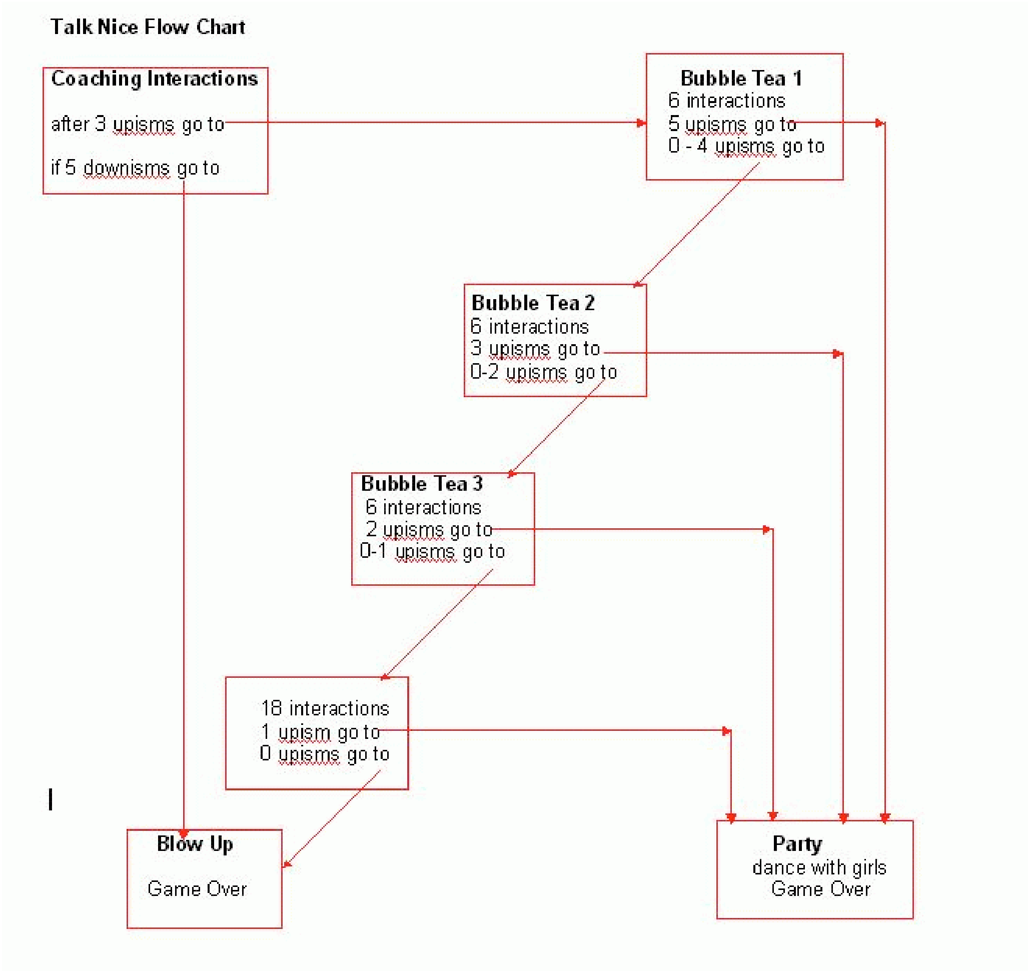


Figure 29.2. Talk Nice flowchart

understand the content of what the user says. The video directs the exchange with relentless cheeriness, setting an agenda of extroverted chat. The girls seem to lead participants into the situation by means of a reward promised at the end (the Party), but really they are persuading through the old teenager technique of setting themselves up as the in-group, taunting everyone else to try and get in, and threatening humiliation for failure. Issues of selfhood do hold sway here, even if there is no intelligent agent that overtly acquires and displays a sense of self. There are questions suggested about where selfhood resides in the interactive situation, and about the impact of the work on both the artificial and human senses of it. Specifically, there is an obvious appeal to a relinquishing of the viewer's self because she or he experiences no option but to play along.

It was Freud who coined the term "ego" for the consciously motivated aspects of human selfhood that involve will, rationality, values, sociality, etc., and it does tend to be the notion of "ego-self" that we mean by "self" in common parlance. There is another approach to selfhood that may apply closely to human-IA dynamics, which is to remove the notion of self from the Freudian tradition that fixates on intrapsychic phenomena, and locate it equally or even predominantly within social relations. In her analysis of human willingness to abandon self in relationships of domination and submission to authority, feminist theorist and psychoanalyst Jessica Benjamin rejects the primacy of the oedipal quest for a lost original unity in the self, and focuses instead on dynamics between self and other that begin in infancy and continue to evolve in adulthood. For Benjamin, domination and submission are signs of failure in the mutuality of recognition within primary relationships that is necessary for a fully realized sense of self. She says, "The need of the self for the other is paradoxical, because the self is trying to establish himself as an absolute, an independent entity, yet he must recognize the other as like himself in order to *be* recognized by him. He must be able to find himself in the other" [5, p. 32]. Our receptiveness or resistance to the authoritarianism of technologies might also be shaped by these deep-seated developmental processes involving our closest relations.

Freud's corollary idea about those aspects of the human psyche that lie outside ego could be described as a kind of excess of self that is outside rational understanding. In my personal absorption of the Freudian schema, there is a "good" excess of self that is fundamentally creative – instinctual, emotional, libidinal, etc. (the "bad" excess of self is a distortion into loss of will or submission to values that have no creative dimension). In George Bataille's writings on the erotic, selfhood or individuation is a trauma of discontinuity with the universe, a splitting from a once unified state that the self is always seeking to repair, an idea closely related to Freud's death instinct. Bataille calls the super-abundance of energy that typifies individuation a plethora, which is always poised for crisis: the cell splitting, or the organism sexually climaxing. The crisis only momentarily resolves the violence of excess energy: ego-self equals ongoing violence and crisis [3, pp. 94-108]. This portrait of too much self I think is closely linked with the Cartesian mind-body split. It is an alternate way of describing a deeply felt ineffectuality in separating the rational mind from the affective domain to reconcile desires, needs and the rest of the human range of experience.

The expansion of the human sensorium that is invoked in multi-sensory art works do exceed the constraints of ego boundaries by appealing directly to affect through senses other than the visual. Consideration of emotion is also one of the more enticing and challenging aspects of modeling social intelligence in autonomous agents. In a territory in-between the two, *Talk Nice* is designed to touch emotional chords as an implicit factor in language exchange. But I don't think that the emotional tone in the work is a direct effect of the characters or the narrative scenarization in the work. Rather, it is an emergent effect. The spontaneous letting go of one's ego-self as an excess of self that submits to the rationalized authority of technology allows for a subsequent re-admitting of emotional response. Ultimately, this signals a re-integration of mind and body. Artworks, IAs and IA-like artifacts can invoke if not a return to oneness with the universe then at least a sense of selfhood and agency shared among humans and our technological objects.

(Editor's note: I think one main difference between embodied art and IA is that the people doing IA have very limited ideas of the experience of users. They are usually overwhelmed with the technical problems of getting anything to work at all. Also, people interacting with IA systems are having very limited experience of an experimental rig, which is a lot different from a daily use of a software product which they have got used to.)

References

[1] Claus Emmeche. *The Garden in the Machine: The Emerging Science of Artificial Life*. Princeton University Press, Princeton, 1994. trans. Steven Sampson.

[2] Jennifer Fisher. Char Davies. *Parachute*, 94:53–54, 1999.

[3] Georges Bataille. *Erotism: Death and Sensuality*. City Lights Books, San Francisco, 1999. trans. Mary Dalwood, first published as *L'Erotisme* in 1957.

[4] Jeanne Randolph. *Psychoanalysis and Synchronized Swimming*. YYZ Books, Toronto, 1991. This is psychiatrist and cultural critic Randolph's set of theoretical essays that delve into the possible subject-object relations between audience and artwork, in the context of Object Relations theory.

[5] Jessica Benjamin. *The Bonds of Love: Psychoanalysis, Feminism, and the Problem of Domination*. Pantheon Books, New York, 1988. Benjamin proposes an "intersubjective view" of self, noting that the concept of intersubjectivity has its origins in the social theory of Jürgen Habermas, encompassing both the individual and social domains, cf. note p. 19.

[6] Simon Penny. The Virtualization of Art Practice: Body Knowledge and the Engineering Worldview. *Art Journal*, 56(3), 1997.

Chapter 30

MULTI-AGENT CONTRACT NEGOTIATION

Knowledge and Computation Complexities

Peyman Faratin
MIT Sloan School of Management

Abstract Two computational decision models are presented for the problem of de-centralized contracting of multi-dimensional services and goods between autonomous agents. The assumption of the models is that agents are bounded in both information and computation. Heuristic and approximate solution techniques from Artificial Intelligence are used for the design of decision mechanism that approach mutual selection of efficient contracts.

1. Introduction

The problem of interest in this chapter is how autonomous computational agents can approach an efficient trading of multi-dimensional services or goods under assumptions of bounded rationality. Trading is assumed to involve negotiation, a resolution mechanism for conflicting *preferences* between selfish agents. We restrict ourselves to a monopolistic economy of two trading agents that meet only once to exchange goods and services. Agents are assumed to be bounded in both information and computation. Information needed for decision making is assumed to be bounded due to both external and internal factors, social and local information respectively. Agents have limited social information because they are assumed to be selfish, sharing little or no information. In addition to this agents may also have limited *local* information (for example over their own preferences) because of complexity of their local task(s). Computation, in turn, is a problem in contract negotiation because of the combinatorics of *scale*. Computation is informally defined as the process of searching a space of possibilities [11]. For a contract with 100 issues and only two alternatives for each issue, the size of the search space is roughly 10^{30} possible contracts, too large to be explored exhaustively.

The unbounded formulation of such an economical problem has long been the central concern of classic game theory which has produced a number of models of social choice. For this reason game theory models have become strong candidates for models of social agents. Surprisingly, such apparently simple games can be used to conceptualize a variety of synthetic, meaningful and formal prototypical context as games. Therefore, such models can be used to design and engineer multi-agent systems as well as analyze the behaviour of the resulting social artifact using the logical tools of the models. However, the underlying unbounded assumptions of classic game theory is problematic for the design of computational systems [2].

Artificial Intelligence (AI) on the other hand has long considered models of the relationship between knowledge, computation and the quality of solution (henceforth referred to as the K-C-Q relationship) [7]. AI has shown that there exists a hierarchy of tradeoffs between K, C and Q, with models that achieve perfect optimal results (like game theory models) but at the cost of requiring omniscience and unbounded agents, to models that sacrifice optimality of Q for a more realistic set of requirements over K and C [12]. Different agent architectures are then entailed from different K-C-Q relationship theories.

In the next two sections two such computational models of negotiation are proposed, one deductive and the other agent-based simulation, that can be analyzed as two different games. The aim of these models has been to attempt to address some of the computational and knowledge problems mentioned above. In particular, in the first model the types of problems of interest is when K is limited because agents have at best imperfect and at worst no knowledge of the others' utility functions. The best an agent can do is to reason with imperfect knowledge by forming approximations of others' utilities. In the second model the knowledge problem is even more extensive because agents in addition are assumed to have an incomplete knowledge of their *own* utility functions.

2. A Bargaining Game

In this model there are two players (a and b) representing one consumer and one producer of a service or a good. The goal of the two agents is to negotiate an outcome $x \in X$, where X is the set of possible contracts describing multi-dimensional goods/services such as the price of the service, the time at which it is required, the quality of the delivered service and the penalty to be paid for reneging on the agreement. If they reach an agreement, then they each receive a payoff dictated by their utility function, defined as $U_i : X \rightarrow [0,1], i \in \{a,b\}$. If the agents fail to reach any deal, they each receive a conflict payoff c. However, from the set X, only a subset of outcomes are "reachable". Call the set of *feasible outcomes* B, containing those agreements that are *individually rational* and bounded by the *pareto optimal* line [13]. An agreement is individually rational if it assigns each agent a utility that is at least as large as the agent can guarantee for itself from the conflict outcome $\mathbf{x_c}$. Pareto optimality is informally defined as the set of outcomes that are better for *both* agents [1]. It is often used as a measure of the efficiency of the social outcome. Given the game $(B, \mathbf{x_c})$, the protocol, or *"rules of encounter"* [8], normatively specifies the *process* of negotiation. The protocol chosen for this game is the alternating sequential model in which the agents take turns to make offers and counter offers [10]. The protocol terminates when the agents come to an agreement or time limits are reached or, alternatively, when one of the agents withdraws from the negotiation. This distributed, iterative and finite protocol was selected because it is un-mediated, supports belief update and places time bounds on the computational resources that can be utilized.

However, like chess for example, agents can have different negotiation strategies given the normative rules of the game. Two heuristic distributed and autonomous search strategies have been developed whose design has been motivated by the knowledge and computation boundedness arguments given above. One parametric mechanism, the *responsive mechanism*, is a mechanism that conditions the decisions of the agent directly to its environment such as the concessionary behaviour of the other party, the time elapsed in negotiation, the resources used, etc. [3]. However, the mechanism is known to have several limitations [4]. In some cases agents fail to make agreements, even though there are potential solutions, because they fail to explore different possible value combinations for the negotiation issues. For instance, a contract may exist in which the service consumer offers to pay a higher price for a service if it is delivered sooner. This contract may be of equal value to the consumer as one that has a lower price and is delivered later. However from the service provider's point of view, the former may be acceptable and the latter may not. The responsive mechanism does not allow the agents to explore for such possibilities because it treats each issue independently and only allows agents to concede on issues.

A second mechanism, called the *trade-off mechanism*, was developed to address the above limitations and consequently select solutions that lie closer to the pareto-optimal line, again in the presence of limited knowledge and computational boundedness [4]. Intuitively, a trade-off is where one party lowers its utility on some negotiation issues and simultaneously demands more on others while maintaining a constant overall contract utility. This, in turn, should make agreement more likely and increase the efficiency of the contracts. An algorithm has been developed that enables agents to make trade-offs between both quantitative and qualitative negotiation issues, in the presence of information uncertainty and resource boundedness for multi-dimensional goods [4]. The algorithm computes n dimensional trade-offs using techniques from fuzzy sim-

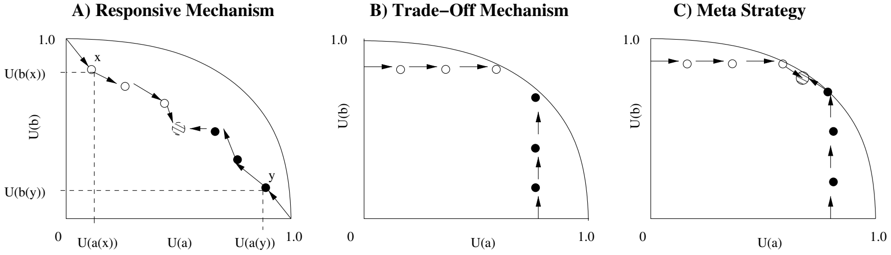


Figure 30.1. Utility Dynamics of the Mechanisms

ilarity [14] to approximate the preference structure of the negotiation opponent. It then uses a hill-climbing technique to explore the space of possible contract trade-offs for a contract that is most likely to be acceptable. The complexity of this algorithm has been shown to grow linearly with growing numbers of issues [4].

The details of the algorithms can be found in [3] and [4]. The dynamics of the contract utility generated by each of the above mechanisms and one possible combination is given in figure 30.1 A, B and C respectively for the alternating sequential protocol. The filled ovals are the utility of the offered contracts from agent a to agent b from agent a's perspective, and the unfilled ovals represent the utility of the offered contracts from agent b to agent a from agent b's perspective. The patterned oval represents the joint utility of the final outcomes. The pareto-optimal line is given by the curvilinear line connecting the two pairs of payoffs $(1, 0)$ and $(0, 1)$. Figure 30.1 A represents *a* possible execution trace where both agents generate contracts with the responsive mechanism. Each offer has lower utility for the agent who makes the offer, but relatively more utility for the other. This process continues until one of the agents is satisfied $(U^a(x^t_{b\to a}) \geq U^a(x^{t'}_{a\to b}))$, where $x^t_{b\to a}$ is the contract offered by agent b to a at time t. This termination criteria is referred to as the cross-over in utilities. The responsive mechanism can select different outcomes based on the rate of concession adopted for each issue (the angle of approach to the outcome point in figure 30.1 A).

Figure 30.1 B represents another possible utility execution trace where both agents now generate contracts with the trade-off mechanism. Now each offer has the same utility for the agent who makes the offer, but relatively more utility for the other (movement towards the *pareto-optimal* line). The trade-off mechanism searches for outcomes that are of the same utility to the agent, but which *may* result in a higher utility for the opponent. Once again, this is a simplification for purposes of the exposition—an offer generated by agent a

may indeed have decreasing utility to agent *b* (arrow moving *away* from the *pareto-optimal* line) if the similarity function being used does not correctly induce the preferences of the other agent.

Finally, agents can combine the two mechanisms through a meta-strategy (figure 30.1 C). One rationale for the use of a meta-strategy is reasoning about the costs and benefits of different search mechanisms. Another rationale, observable from the example shown in figure 30.1 B, is that because the local utility information is private agents can not make an interpersonal comparison of individual utilities in order to compute whether a pareto optimal solution has indeed been reached. In the absence of a mediator the lack of such global information means negotiation will fail to find a joint solution that is acceptable to both parties. In fact agents enter a loop of exchanging the same contract with one another. Figure 30.1 C shows a solution where both agents implement a responsive mechanism and concede utility. This concession may, as shown in figure 30.1 C, indeed satisfy the termination conditions of the trade-off mechanism where offers cross-over in utilities. Alternatively, agents may resume implementing a trade-off algorithm until such a cross-over is eventually reached or time limits are reached. In general, the evaluation of which search should be implemented is delegated to a meta-level reasoner whose decisions can be based on bounding factors such as the opponent's perceived strategy, the on-line cost of communication, the off-line cost of the search algorithm, the structure of the problem or the optimality of the search mechanism in terms of completeness (finding an agreement when one exists), the time and space complexity of the search mechanism, and the expected solution optimality of the mechanism when more than one agreement is feasible.

3. A Mediated Game

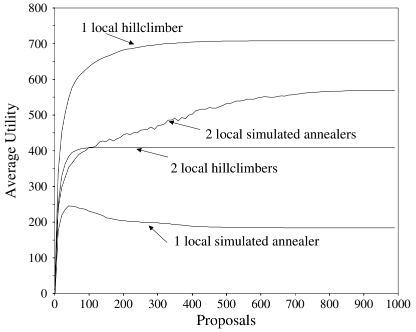
In the above model the issues being negotiated over are assumed to be independent, where the utility to an agent of a given issue choice is independent of what selections are made for other issues. The utility function that aggregates the individual utilities under this assumption is then taken to be linear. This assumption significantly simplifies the agents' local decision problem of what issue values to propose in order to optimize their local utility. Optimization of such a linear function is achieved by hillclimbing the utility gradient. However, real world contracts, are highly inter-dependent. When issue interdependencies exist, the utility function for the agents exhibits multiple local optima. Multi-optimality results in firstly a more extensive bounded rationality problem since not only is computation limited but now also both *local* and global knowledge are limited. Local knowledge is limited because the agent now has to know and optimize a much more complicated utility function. Secondly, a methodological change from deductive models to simulation studies is needed due to the complex non-linearities involved in the system. The solution to these problems are briefly outlined below in a model of negotiation that departs from the more deductive model outlined above [5].

In this model a contract x is an N dimensional boolean vector where $x_i \in \{-1, +1\}$, represents the presence or absence of a "contract clause" i. The contract search policy is encoded in the negotiation protocol. Because generating contract proposals locally is both knowledge and computationally expensive we adopt an indirect single text protocol between two agents by delegating the contract generation process to a centralized mediator [9]. A mediator proposes a contract x^t at time t. Each agent then votes to accept or reject x^t. If both vote to accept, the mediator iteratively mutates the contract x^t and generates x^{t+1}. If one or both agents vote to reject, a mutation of the most recent mutually acceptable contract is proposed instead. The process is continued until the utility values for both agents become stable (i.e. until none of the newly contract proposals offer any improvement in utility values for either agent). Note that this approach can straightforwardly be extended to N party (i.e. multi-lateral) negotiation. The utility of the contract to an agent is defined as the linear combination of all the pairwise influences between issues.

Two computationally inexpensive decision algorithms were evaluated in this protocol: a hillclimber and a simulated annealer . A hillclimber only accepts a contract if and only if the utility of the contract x increases monotonically when *an* issue is changed. However, this steepest ascend algorithm is known to be incapable of escaping local maxima of the utility function. The other decision algorithm is based on the knowledge that search success can be improved by adding thermal noise to this decision rule [6]. The policy of decreasing T with time is called simulated annealing [6]. Simulated annealing rule is known to reach utility equilibrium states when each issue is changed with a finite probability and time delays are negligible.

To evaluate these algorithms simulations were run again with two agents a and b. The contract length N was set to 100 (corresponding to a space of 2^{100}, or roughly 10^{30} possible contracts) where each bit was initialized to a value $\{-1, +1\}$ randomly with a uniform distribution. The initial temperature was set to 10 and decreased in steps of 0.1 to 0. Final average utilities were collected for 100 runs for each temperature decrement.

The left figure in figure 30.2 shows the observed individual payoffs for tests examining the relationship of C-Q with local utility metric of Q. One observation is that if the other agent is a local hill-climber, an agent is then individually better off being a local hill-climber, but fares very badly as local annealer. If the other agent is an annealer, the agent fares well as an annealer but does even better as a hillclimber. The highest *social* welfare, however, is achieved when both agents are annealers. This pattern can be readily understood as follows. At high virtual temperature an annealer accepts almost all proposed contracts in-



	Annealer	Hillclimber
Annealer	550/550	180/700
Hillclimber	700/180	400/400

Figure 30.2. Game Dynamics (left) and Final Payoff Matrix of the Game (right)

dependently of the cost-benefit margins. Therefore, at high virtual temperature the simulated annealer is more explorative and "far sighted" because it assumes costs now are offset by gains later. This is in contrast to the myopic nature of the hillclimber where exploration is constrained by the monotonicity requirement. In the asymmetric interaction the cooperation of annealers *permits* more exploration of the contract space, and hence arrival to higher optima, of hillclimber's utility landscape. However, this cooperation is not reciprocated by hillclimbers who act selfishly. Therefore, gains of hillclimbers are achieved at the cost of the annealer. The right figure in figure 30.2 represents the underlying game as a matrix of final observed utilities for all the pairings of hillclimber and annealer strategies. The results confirm that this game is an instance of the prisoner's dilemma game [1], where for each agent the dominant strategy is hillclimbing. Therefore, the unique dominating strategy is for both agents to hillclimb. However, this unique dominating strategy is pareto-optimally dominated when both are annealers. In other words, the single Nash equilibria of this game (two hillclimbers) is the only solution not in the Pareto set.

4. Conclusions

The contracting problem was used to motivate two different heuristic and approximate agent decision models, both based on a realistic set of requirements over both K and C. However, the cost of these requirements is the sub-optimality of Q. This trade-off was demonstrated in both models by negotiation strategies selecting outcomes that are not pareto efficient. However, imperfections is a common feature of the world and real social systems have established personal and institutional mechanisms for dealing with such imperfections. Similarly, in future computational models are sought that are incremental, repeated and support feedback and error-correction. Learning and evolutionary techniques are two candidates for optimizing this trade-off given the environment of the agent.

References

[1] K. Binmore. *Fun and Games: A Text on Game Theory* Lexington, Massachusetts: D.C. Heath and Company, 1992.

[2] P. Faratin. *Automated Service Negotiation between Autonomous Computational Agents* Ph.D. Thesis, Department of Electronic Engineering, Queen Mary and Westfield College, University of London, 2000.

[3] Faratin, P and Sierra, C and Jennings, N.R. *Negotiation Decision Functions for Autonomous Agents* , Journal of Robotics and Autonomous Systems, 24(3–4):159–182, 1998.

[4] Faratin, P and Sierra, C and Jennings, N.R. *Using Similarity Criteria to Make Negotiation Trade-Offs* , International Conference on Multi-agent Systems (ICMAS-2000), Boston, MA., 119-126, 1998.

[5] P. Faratin, M. Klein, H. Samaya and Yaneer Bar-Yam. *Simple Negotiating Agents in Complex Games: Emergent Equilibria and Dominance of Strategies* , In Proceedings of the 8th Int Workshop on Agent Theories, Architectures and Languages (ATAL-01), Seattle, USA, 2001.

[6] S. Kirkpatrick and C.D. Gelatt and M.P. Vecci. *Optimization by Simulated Annealing* , Science 671–680, 1983.

[7] J. Pearl. *Heuristics* Reading, MA: Addison-Wesley, 1984.

[8] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter* Cambridge, Massachusetts: MIT Press, 1994.

[9] H. Raiffa. *The Art and Science of Negotiation* , Cambridge, MA: Harvard University Press, 1982.

[10] A. Rubinstein. *Perfect equilibrium in a bargaining model* , Econometrica, 50:97–109, 1982.

[11] S. Russell and P. Norvig. *Artificial Intelligence: A modern approach* Upper Saddle River, New Jersey: Prentice Hall, 1995.

[12] S. Russell and E. Wefald. *Do the Right Thing* Cambridge, Massachusetts: MIT Press, 1991.

[13] J. von Neumann and O. Morgernstern. *The Theory of Games and Economic Behaviour* Princeton, N.J: Princeton University Press, 1944.

[14] L. A. Zadeh. *Similarity relations and fuzzy orderings* Information Sciences, 3:177–200, 1971.

Chapter 31

CHALLENGES IN AGENT BASED SOCIAL SIMULATION OF MULTILATERAL NEGOTIATION

Scott Moss
Centre for Policy Modelling, Manchester Metropolitan Univeristy

Abstract This paper is an interim report on the development of an analysis of negotiating positions and strategies in a complex environmental management situation. There are seven categories of negotiating parties with many issues to be resolved. Each issue could be resolved in a large number of ways. An abstract model that captures the structure of the negotiations is reported. Simulations suggest that, while bilateral negotiations readily reach agreement, multilateral negotiations do not. The way forward for both modelling a the design of negotiation procedures will require historical evidence about successful multilateral negotiations.

1. Introduction

It is not hard to find examples of failed negotiations. Recently, we have seen the failure of attempts to build on the Kyoto agreement on reducing green house gas emissions, the breakdown of the Oslo Accord under which Israel and Palestine were moving towards a peaceful settlement of their differences, the failure of OECD members to agree on trade liberalisation measures, the halting progress of the Northern Ireland settlement under the terms of the Good Friday Agreement.

At the same time, there are clearly many examples of successful negotiation that form part of the small change of everyday life. In many households, partners easily agree on what they shall have for dinner or how they shall spend the evening or weekend. More momentously, couples agree to marry or cohabit. The negotiations of transactions in houses are sometimes difficult but frequently resolved. Even the distribution of assets in the course of divorce proceedings is regularly achieved by agreement between the partners to the dissolving marriage.

It is clear that the examples of difficult negotiations involve both more parties and larger numbers of related issues than do the examples of regularly successful negotiations. But there is a second difference, as well. The examples of success are negotiations among two parties and if the parties are in fact composed of several individuals, within each party there are no differences of goals. Whereas the large scale negotiations generally have to reconcile a wide range of interests. In Northern Ireland, there are ranges of both Loyalist and Nationalist groups and there are frequently violent incidents among such groups within the same sectarian community. An analogous description would be apposite to the Israeli-Palestinian or many other difficult, apparently bilateral, negotiations.

This paper is an interim report on the development of techniques for modelling multilateral negotiation. To model bilateral negotiation turns out be very straightforward but, though the modelling framework was set up to extend easily to represent negotiation among any number of parties, it is extraordinarily difficult to capture the process of convergence of positions among three or more parties. The nature of the difficulties encountered suggest that models of failed negotiations provide insights into the reasons why difficulties are encountered in real social processes. A promising means of learning about processes of successful multilateral negotiations is to describe real instances of successful multilateral negotiations with agent based social simulation models.

An elaboration of this suggestion is presented in the concluding section 5 on the basis of the model described in some detail in section 3, the results of the model with two and then with more than two negotiating agents is presented in section 4.

2. A Model Of Multi Lateral Negotiation

The model reported here is the prototype for a description of stakeholder negotiation in the Limberg basis of the River Meuse. There are seven such stakeholders and a large number of issues to be resolved.

The stakeholders are ministries of the Netherlands national government, the provincial government of Limberg, farmers, NGOs (mainly concerned with the creation of nature reserves), shipping companies, gravel extraction companies, households and community organisations. The issues being negotiated include flood control, navigation, gravel extraction, the creation and maintenance of nature reserves, agriculture. There are manifold - certainly more than two - outcomes for many of the individual negotiating issues. Consequently, any suitable representation of the negotiating process has to take into account the multiplicity of stakeholders, issues and outcomes for each issue.

Over the past decade, there have been several plans with changing objectives for the Meuse. The structure of these plans, and the relative importance of their objectives, has changed with each of two major floods in the 1990s. After each flood, the importance of flood control and population safety became - for a time - more dominant. Also, individual plans for navigation, flood control and other issues were integrated eventually into a single plan under the aegis of the Maasverkenprojekt. On no occasion has full agreement been reached among all of the negotiating parties.

The first model reported here does not describe the actual issues but instead represents the structure of the issues involved. Successive models will incorporate the issues with increasing explicitness and no model will distort the issues or relations among the negotiators "for the sake of simplicity".

2.1 Abstract representation of agents' positions

The negotiating stance of each agent is represented by two digit strings. One string - the agent's *position string* - represents the preferred outcome of the negotiating process with respect to each issue under discussion. The other string - the agent's *importance string* - represents the importance the agent attaches to achieving its preferred outcome for each issue. For example, and agent's desired outcomes might be represented by the position string

[2 1 4 2 3 0 0 3 2 4 1 0 2 1]

where the value at each index of the string is a representation of the desired outcome of the negotiating process for a particular issue. The issue corresponding to each index of the position string is the same for every agent. The number of integer values that can be assigned to any position is determined by the model operator at the start of each simulation run with the model. In this case, the values taken at each index of the position string are in the interval [0,4].

The corresponding importance string of the agent might be

[3 1 0 2 0 3 3 1 0 2 3 1 0 1]

indicating that the most important objectives of the agent (indicated by the 3s in the importance string) are to obtain a value of 2 for the issue denoted by the first digit of the strings and the value 0 for the sixth and seventh issues and the value 1 for the 11th issue.

The effect of the negotiation process is necessarily represented as changes in the position strings of the participating agents. Moreover, although not implemented in the simulations reported below, it seems likely that the importance attached to different positions will also change over the course of the negotiation process - perhaps as it becomes important to maintain common positions important to partners which whom agreement has been reached.

2.2 Selection of negotiating partners

Agents could have any of a wide variety of strategies for the identification of issues about which to negotiate and for the selection of negotiating partners. At one extreme, an agent could identify an issue and then negotiate with every possible (or known) agent concerning that issue. At the other extreme, agents can select other agents with which to negotiate and determine the issues in collaboration with the selected agents. The strategy to be modelled - whether one of these extreme cases or some combination or set of parallel strategies - should depend on observation and the evidence of domain expertise.

In the model reported here, the negotiating strategy was driven by the selection of agents as negotiating partners. The criteria for selecting an agent with which to negotiation were based on trustworthiness, reliability, similarity, helpfulness, acquaintanceship, untrustworthiness, unreliability, unhelpfulness. One agent identifies another as reliable if the other agent responds affirmatively to a suggestion that the two agents negotiate. An agent will identify another as trustworthy if its public negotiating position reflects previous agreements between the two agents. An agent is helpful if it suggests to two or more other agents that they might usefully negotiate with one another and agreement among those agents is realised. An agent will identify another as similar if, among all of the negotiating positions known to the agent, the other agent shares the largest number of position values. One agent can know another either because of an approach at random or because the other agent has made contact by suggesting a negotiation.

Each agent in the model has rules for attaching *endorsements* - tokens reflecting the selection or aversion criteria - to other agents. The ranking of the importance of endorsements is, in the first instance, random except that opposite endorsements (helpful and unhelpful, trustworthy and untrustworthy, reliable and unreliable) have rankings of the same magnitude and opposite sign. So that if trustworthy is the most important positive endorsement, untrustworthy will be the most important negative endorsement. Each agent will have its own initial ranking of positive (and therefore negative) endorsements. Each agent will select the best endorsed agent it knows as a negotiating partner at each stage.

Over the course of a negotiation process, each agent will continue to learn about other agents - a process represented by the ongoing attachment of endorsements. Each agent also learns which are the most important criteria to use in selecting negotiating partners. If the use of a particular set of rankings of criteria leads to agreement with a selected agent or group of agents, there is no reason to change the relative importance of the different criteria. If no agreement is reached, then there will be less confidence in the current ranking -

though it is unlikely that a wholesale change in rankings will follow from every failure to achieve some agreement.

In order to capture this learning process about endorsements and their relative values, agents' learning is represented by the Chialvo-Bak [1] algorithm. This algorithm uses a sort of neural network approach but without positive reinforcement of synapse weights. In the present case, the input neurons are attached to endorsement tokens and the output neurons are ranking values to be attached to the endorsements. There were five intermediate layers, each containing 40 neurons. Starting with the input neurons, each neuron as seven synapses out to the next layer until the output neuron layer is reached. The paths followed from input to output neurons is determined by the synapse with the highest weight emanating from each neuron. When agreement is not reached, the value of each synapse on the dominant path is reduced by a small amount (usually by one per cent) and the sum of the reductions is distributed equally among the rest of the (2000+) synapses. Consequently, changes in the behaviour of an agent take place relatively infrequently but will, from time to time, be fairly extensive.

There are two advantages to be gained from implementing this learning process. One is that the simulations determine the most important criteria to be used in choosing negotiating partners. The other is the flexibility of the ordering of criteria since it is possible that the importance of different criteria will change over the course of any negotiation process. It is possible, for example, that reliability is most important at early stages so that there is some meaningful communication but that trustworthiness is most important in the final stages.

2.3 Negotiation strategy

It is a commonplace in the negotiation literature that the least important issues should be addressed first. Once negotiating styles have accommodated one another and a recognition of reliability and trustworthiness established, there is a basis for considering more important substantive issues. The most difficult issues are left to the last.

Every agent in the model reported here adopts this sort of strategy. Each agent offers to its preferred negotiating partner a list of positions for the issues the agent found least important among all of the issues that had not yet been resolved. Denote the first agent as A and A's preferred negotiating partner as P. If P made some offer of negotiating positions then, if that offer contained values for positions that A found least important, and also some values that A found to be more important, then A would accept P's offer on the least important issues in exchange for P's acceptance of the same number of A's positions. In general terms, some agreement could always be reached provided the two

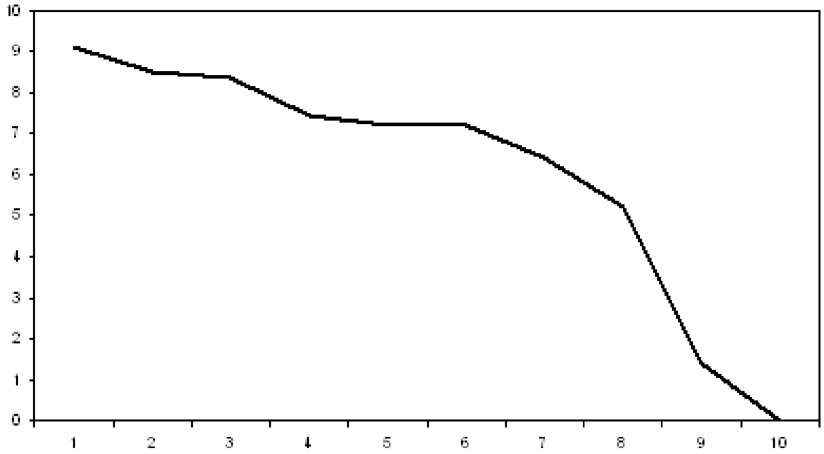


Figure 31.1. Distance between 2 agents in bilateral negotiation

agents preferred to negotiate with one another and each was able to offer to change one or more of its least important positions in exchange for the other agent agreeing one of its more important positions.

Once any pair or larger group of agents fully agrees on all positions, they form a coalition to negotiate with agents not in the coalition or with other coalitions. The process ends when all agents are members of a single coalition or super-coalition (i.e. coalition of coalitions of coalitions ...). In practice, the only simulated negotiation processes that reached a conclusion were all of the two-agent processes.

3. Simulation Results

The progress of bilateral negotiation was represented by changes in the differences of negotiating positions of two agents. These differences were measured as the Euclidian distance between the two position strings interpreted as coordinate vectors in a 30-dimensional hyperspace. An example of the progress represented by this measure is given in Figure 31.2. This progress is typical of all runs with two negotiating agents. The range of the number of cycles elapsed before agreement was reached was from 8 to 12 with the bulk of the distance eliminated in the last half or less of the cycles. There was no learning for the agents to do since they had no choice of negotiating partners.

Although simple negotiating strategies work well for the modelled bilateral negotiation, they do not work at all in simulations of multilateral negotiation

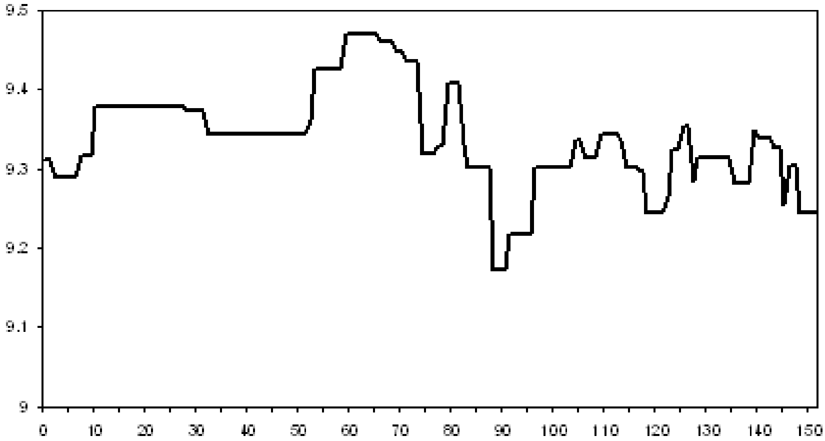


Figure 31.2. Average distance between negotiating positions of agents in nine-agent simulation

with three or more agents. Simply trading agreements on more important positions in exchange for giving up less important positions is evidently insufficient. The problem here is that moving towards agreement with any other agent typically involves increasing the distance to some other agent. It is no doubt possible to devise a variety of arrangements under which agents combine in pairs to reach agreement and form a coalition and then pairs of coalitions negotiate to form a super-coalition and so on until every agent is in the coalition. The value of such an exercise is not clear. Certainly there is no evidence that such a tree of bilateral agreements is a realistic description of successful negotiations, though equally certainly there is some element of small groups coming together on particular issues.

4. Implications

If good science starts from good observation, then the implications of these simulation results are that we should model actual, individual processes of multilateral negotiation. The modelling itself will doubtless yield insights into the elements of successful and unsuccessful negotiation processes and the modelling of a range of such processes is likely to inform the development of modelling techniques that apply quite generally to descriptive simulation models and to capture sound negotiating processes that will usefully inform the development of multi agent software systems.

The results reported above indicate that it will be much more difficult to simulate successful negotiations among three or more agents and, therefore, much more difficult to design software agents and mechanisms for general multi lateral negotiation. In general, bilateral negotiation is a special case and there is no reason to infer anything about negotiations among three or more parties from results with models of bilateral negotiation.

The decision by each agent concerning which other agents to engage in negotiation is far from trivial. In the model, agents were concerned with the trustworthiness, reliability, helpfulness and similarity of other agents. Agents did not appear to learn which, if any, of these characteristics should be given priority in selecting negotiating and coalition partners.

In general, it would be hard to justify as good science the repeated revision of abstract simulation models until we found one that produced convergence in a negotiating process and then to assert that such a model describes a socially useful approach to negotiation. Producing such a model is a purely intellectual exercise. To be useful, it must be validated. To be validated, it must be shown to be a good descriptor of actual successful multi lateral negotiations. If such a model can be validated against a range of negotiating processes, we might then have some confidence in the model as a pattern for good negotiating practice. It is hard to see any substantive difference between validating abstract models and building models around descriptions of actual negotiations. Both involve the development of a general understanding by means of the development of descriptively accurate simulation models.

References

[1] Chialvo, D.R. and Bak P. Learning from Mistakes, *Neuroscience* 90:1137-1148, 1999.

Chapter 32

ENABLING OPEN AGENT INSTITUTIONS

Juan A. Rodríguez-Aguilar[1] and Carles Sierra[2]

[1] *iSOCO (Intelligent Software Components)*

[2] *Institut d'Investigació en Intel.ligència Artificial, Spanish Scientific Research Council (CSIC)*

Abstract In this paper we argue that open multi-agent systems can be effectively designed and implemented as *electronic institutions* composed of a vast number of heterogeneous (human and software) agents playing different roles and interacting by means of speech acts. Thus taking inspiration from traditional human institutions, we offer a general agent-mediated computational model of institutions that serves to realise an actual agent-mediated electronic auction house where heterogeneous agents can trade.

1. Introduction

Up to date most of the work produced by multi-agent systems(MAS) research has focused on systems developed and enacted under centralised control. Thus, MAS researchers have bargained for well-behaved agents immersed in reliable infrastructures in relatively simple domains. Such assumptions are not valid when considering *open systems* [3] whose components are unknown beforehand, can change over time and can be both human and software agents developed by different parties. Examples of open agent systems include open electronic marketplaces and virtual supply chains, disaster recovery operations, collaborative design and international coalition military forces.

Although open systems have recently started to be considered by MAS researchers as one the most important application of multi-agent systems, their inherent issues (agent heterogeneity, reliability, accountability, legitimacy, societal change, etc.) have not been conveniently addressed yet. And then how to approach their design and construction? Although there has been a surge of interest in agent-oriented methodologies and modelling techniques in the last few years motivated and spurred by the first generation of agent developments [7, 8, 11], at present most agent applications lack a principled methodology underpinning their development, and so they are produced in an ad hoc fashion.

Another fundamental aspect is to opt for either a micro (agent-centered) view or a macro (organisation-centered) view of MAS. Although early work in DAI identified the advantages of organisational structuring as one of the main issues in order to cope with the complexity inherent to designing DAI systems (f.i.[2]) MAS research has traditionally kept an individualistic character, evolving patterned on a strong agent-centered flavour. And yet, there is an increasing interest in incorporating organisational concepts into MAS as well as in shifting from agent-centered to organisation-centered designs [1, 5, 7] that consider the organisation as a first-class citizen. Nonetheless, in general the introduction of social concepts into multi-agent systems has been undertaken in a rather informal way.

In this paper we adopt a macro perspective in order to effectively construct open multi-agent systems. Thus we argue on the need for deploying normative environments similar to those provided by human institutions following the pioneering work in [5]. Institutions [6] represent the rules of the game in a society, including any (formal or informal) form of constraint that human beings devise to shape human interaction. They are the framework within which human interaction takes place, defining what individuals are forbidden and permitted and under what conditions. Furthermore, institutions are responsible for ascertaining violations and the severity of the punishment to be enacted. We uphold that open multi-agent systems can be successfully designed and implemented as institutionalised agent organisations (henceforth *electronic institutions*).

In Section 2 we present a case study of human institution in order to subsequently identify its components in Section 3. Next, in Section 4 we describe the two types of agents on which we found a computational model of electronic institution which successfully served to realise an actual agent-mediated electronic auction house. Finally, Section 5 contains some conclusions.

2. The Fish Market. An Actual-world Human Institution

As a starting point for the study of institutions we choose the fish market as a paradigm of traditional human institutions. The actual fish market can be described as a place where several *scenes* take place simultaneously, at different places, but with some causal continuity. Each scene involves various agents who at that moment perform well-defined functions. These agents are subject to the accepted market conventions, but they also have to adapt to whatever has happened and is happening at the auction house at that time. The principal scene is the auction itself, in which buyers bid for boxes of fish that are presented by an auctioneer who calls prices in descending order —the downward bidding protocol. However, before those boxes of fish may be sold, fishermen have to deliver the fish to the fish market (in the *sellers' admission scene*) and buyers need to register for the market (at the *buyers' admission scene*). Likewise, once a box of fish is sold, the buyer should take it away by passing through a *buyers' settlements scene*, while sellers may collect their payments at the *sellers' settlements scene* once their lot has been sold.

3. Institution Components

In order to engineer open agent multi-agent systems as electronic institutions we must firstly identify the core notions and components of electronic institutions, the computational counterpart of institutions, taking inspiration on the case study presented above. Thus our conception of electronic institution shall be founded on the following concepts:

Agents and Roles. Agents are the players in an electronic institution, interacting by the exchange of illocutions (speech acts), whereas roles are standardised patterns of behaviour. Any agent within an electronic institution is required to adopt some role(s). We fundamentally distinguish two classes of roles: *institutional*, and *non-institutional.*

Dialogical framework. In a dialogical institution, agents interact through illocutions. Institutions establish the ontology and the common language for communication and knowledge representation, which are bundled in what we call dialogical framework. By sharing a dialogical framework, we enable heterogeneous agents to exchange knowledge with other agents.

Scene. Interactions between agents are articulated through agent group meetings, which we call *scenes*, with a well-defined communication protocol. We consider the protocol of a scene to be the specification of the possible dialogues agents may have to articulate a multi-agent activity. A scene defines a role-based framework of interaction for agents. A distinguishing feature of scenes is that agents may join in or leave during the activity.

Performative structure. Scenes can be connected, composing a network of scenes, the so-called *performative structure*, which captures the existing relationships among scenes. A performative structure specifies how agents can legally move from scene to scene by defining both the pre-conditions to join in and leave scenes. Considering the fish market, while some activities like the admission of buyers and sellers are completely independent, others are tightly related. For instance, a buyer cannot bid for any good unless he has previously and successfully set up a credit line.

Normative Rules. Agent actions in the context of an institution have consequences, usually in the shape of compromises which impose obligations or restrictions on dialogic actions of agents in the scenes wherein they are acting or will be acting in the future. For instance, after winning a bidding round the bidder is committed to subsequently pay for the acquired good. Obligations and prohibitions are captured by means of normative rules.

Based on the institution components introduced above, in [8] we offer a formal specification of electronic institutions that founds the computational model presented in Section 4.

4. Agent-mediated Institutions

The workings of an electronic institution can be fully realised by means of the articulation of two types of agents: *institutional agents* and *interagents*. Institutional agents are those to which the institution delegates its services, whereas interagents are a special type of facilitators that mediate all the interactions of external agents within an electronic institution and enforce institutional rules. Our agent-mediated computational model (thoroughly detailed in [8].) has proven its usefulness in the development of FM96.5, the computational counterpart of the fish market [10], which served as the basis for the subsequent development of FM, an agent-mediated test-bed for auction-based markets[9].

4.1 Institutional Agents

An institution delegates part of its tasks to agents adopting institutional roles (in the fish market the auctioneer is responsible for auctioning goods, the sellers' admitter for registering goods, and the accountant for the accounts' book-keeping). We refer to this type of agents as institutional agents. An institutional agent can possibly adopt multiple institutional roles. In order to fully specify an institutional role we must specify its *life-cycle* within an institution in terms of its responsibilities along with the policy of responsibilities' management.

More concretely, we specify an institutional role's life-cycle as a regular expression built by combining the following operations: $x.y$ (*x* followed by *y*), x^* (*x* occurs 0 or more times), $x||y$ (*x* and *y* interleaved), $x|y$ (*x* or *y* occurs), x^+ (*x* occurs 1 or more times), $[x]$ (*x* is optional); where x and y stand for scene (activity) names. Table 32.1 contains the specification of the *buyer_admitter*, *auctioneer* and *seller_accountant* roles in FM96.5, the computational counterpart of the fish market. For instance, an institutional agent playing the *buyer_admitter* role must firstly enter at the *registry* scene with the *boss* of the market. Next, it is expected to meet other institutional agents (auctioneer, buyers' accountant, sellers' admitter and sellers' accountant) at the *opening* scene. Afterwards it can start processing buyers' requests for admission.

Institutional agents might be required to comply with several responsibilities at the same time. In such a case, an institutional agent must know how to prioritise (schedule) simultaneous responsibilities. For this purpose, responsibilities are ranked according to their relevance. As an example, the responsibilities

buyer_admitter	$= registry.((opening.(buyer\_admission)^*)\Vert closing)$
auctioneer	$= registry.((opening.(request\_goods.(auction\Vert credit\_line))^*)$ $\Vert$ closing)
seller_accountant	$= registry.(opening.(good\_adjudication^*\Vert seller\_settlements^*)$ $\Vert$ closing)

Table 32.1. Institutional agents' responsibilities specification.

for the auctioneer are ranked as follows: *(registry,High) (opening,High) (closing,High) (request_goods,Medium) (credit_line,Medium) (auction,Low)*; where *High*, *Medium* and *Low* denote different priority degrees..

And yet there remains the matter of deciding how to behave within each scene in which an institutional agent will get involved. When participating in a scene, at some states an institutional agent will be expected to *act* by uttering an illocution as a result of an inner decision-making process. For instance, an auctioneer must know how to select the winner of a bidding round, a buyers' admitter must decide whether to admit a buyer or not, and a sellers' admitter must know how to tag the incoming goods to be put at auction. These inner activities yield illocutions to be uttered by the institutional agent. Since an institutional agent must know which method to fire at those scene states at which it is expected to act, his *behaviour specification* is provided as a collection of methods to be fired at particular states of the scene. For instance, Figure 32.1 contains a specification of an auction scene protocol (the graph nodes denote scene states connected by arcs labeled by illocution schemes. Transitions occur when illocutions uttered by agents match illocution schemes.). The auctioneer is instructed to run the *declareWinner* method at ω_7.

In [8] we propose a general model of institutional agent in order to ease development. Thus, the very same institutional agent model (architecture) can be employed to deploy several institutional agents playing different roles.

4.2 Interagents

Interagents [4] constitute the sole and exclusive means through which agents interact with the rest of agents within the institution. They become the only channel through which illocutions can pass between external agents and institutional agents. Notice that interagents are all owned by the institution but used by external agents. The mediation of interagents is key in order to guarantee: the legal exchange of illocutions among agents within scenes; the sound transition of external agents from activity to activity within the institution's performative structure; the enforcement of institutional rules; and the accountability of external agents' interactions.

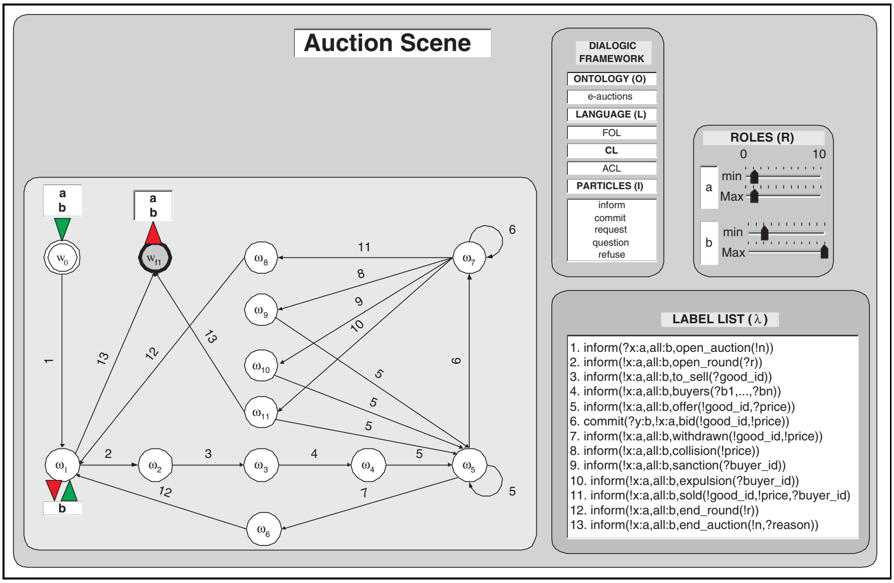

Auction Scene
DIALOGIC FRAMEWORK
ONTOLOGY (O)
e-auctions
LANGUAGE (L)
FOL
CL
ACL
PARTICLES (I)
inform
commit
request
question
refuse
ROLES (R)
0
10
a
min
Max
b
min
Max
LABEL LIST (λ)
1. inform(?x:a,all:b,open_auction(!n))
2. inform(!x:a,all:b,open_round(?r))
3. inform(!x:a,all:b,to_sell(?good_id))
4. inform(!x:a,all:b,buyers(?b1,...,?bn))
5. inform(!x:a,all:b,offer(!good_id,?price))
6. commit(?y:b,!x:a,bid(!good_id,!price))
7. inform(!x:a,all:b,withdrawn(!good_id,!price))
8. inform(!x:a,all:b,collision(!price))
9. inform(!x:a,all:b,sanction(?buyer_id))
10. inform(!x:a,all:b,expulsion(?buyer_id))
11. inform(!x:a,all:b,sold(!good_id,!price,?buyer_id)
12. inform(!x:a,all:b,end_round(!r))
13. inform(!x:a,all:b,end_auction(!n,?reason))
a
b
w0
a
b
wf1
ω1
b
ω2
ω3
ω4
ω5
ω6
ω7
ω8
ω9
ω10
ω11
1
2
3
4
5
5
5
5
5
6
6
7
8
9
10
11
12
12
13
13

Figure 32.1. Graphical Specification of an Auction Scene

One of the fundamental tasks of interagents is to ensure the legal exchange of illocutions among the agents taking part in some scene: what can be said, to whom and when. For this purpose, interagents employ *conversation protocols* (CP) [4]. CPs define coordination patterns that constrain the sequencing of illocutions within a scene and allow to store, and subsequently retrieve, the contextual information (illocutions previously sent or heard) of ongoing scenes. We can think of CPs as scenes extended with the necessary actions to keep contextual information. Based on contextual information, when receiving some illocution from an external agent to be transmitted, an interagent can assess whether the illocution is legal or else whether it must be rejected or some enforcement rule activated.

Consider the auction scene. A buyer agent receives the prices called by the auctioneer through his interagent, which keeps track of the latest price called. When the buyer agent submits a bid, his interagent collects it and verifies whether the buyer is bidding for the latest offer price. If so, the interagent posts the bid to the auctioneer, otherwise it's rejected. Once the bid has been submitted, the buyer is not allowed to re-bid. If he tries, their bids are disallowed, and if he compulsively tries his interagent unplugs him from the institution. Then his interagent autonomously follow the required procedures to log the buyer out from the auction house.

Interagents also constrain external agents' behaviour in their transition between scenes. Figure 32.2 depicts the specification of the performative structure projection for buyer agents in FM96.5, the computational counterpart of the fish market. If some buyer requests his interagent for leaving the institution after making some acquisitions in the auction scene, his interagent will refuse the request because the agent has pending obligations: the payment of the acquired goods, as stated by the institutional normative rules.

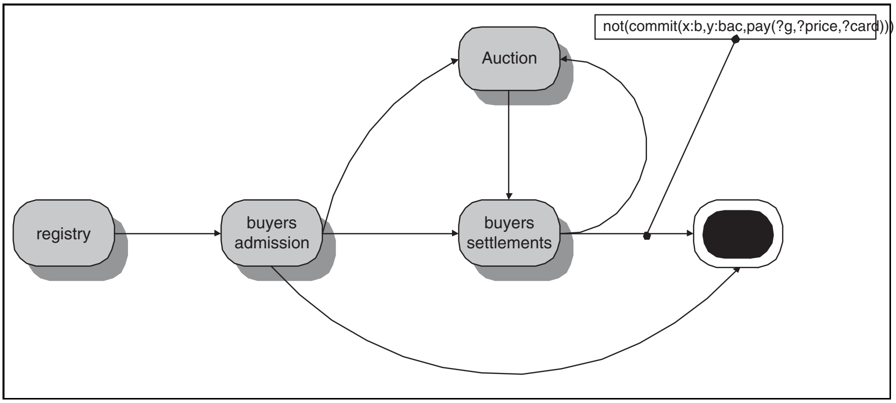


Figure 32.2. Performative structure projection for buying agents.

In general, based on external agents' actions, the facts deriving from their participation in scenes and the institutional normative rules, interagents are capable of determining which obligations and prohibitions to trigger.

Finally, interagents handle transparently to external agents their incorporation into ongoing scenes, their exit from ongoing scenes, their migration between scenes, and the joint creation of new scenes with other agents by means of their coordinated activity with institutional agents, as fully accounted by the computational model detailed in [8].

5. Conclusions

Organisational and social concepts can enormously help reduce the complexity inherent to the deployment of open multi-agent systems. In particular, institutions are tremendously valuable to help solve the many inherent issues to open multi-agent systems. The conception of open multi-agent systems as electronic institutions lead us to a general computational model based on two types of agents: institutional agents and interagents. Although our computational model proved to be valuable in the development of the computational counterpart of the fish market, we claim that such a computational model is general enough to found the development of other agent institutions.

Acknowledgments

This work has been partially funded by research grant FI-PG/96-8.490. This research was performed at IIIA-CSIC.

References

[1] Ferber, J. and Gutknetch, O. A meta-model for the analysis of organizations in multi-agent systems. In *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*, pages 128–135, 1998.

[2] Gasser, L., Braganza, C., and Herman, N. *Distributed Artificial Intelligence*, chapter MACE: A flexible test-bed for distributed AI research, pages 119–152. Pitman Publishers, 1987.

[3] Hewitt, C. Offices are open systems. *ACM Transactions of Office Automation Systems*, 4(3):271–287, 1986.

[4] Martín, F. J., Plaza, E., and Rodríguez-Aguilar, J. A. An infrastructure for agent-based systems: An interagent approach. *International Journal of Intelligent Systems*, 15(3):217–240, 2000.

[5] Noriega, P. *Agent-Mediated Auctions: The Fishmarket Metaphor.* Number 8 in IIIA Monograph Series. Institut d'Investigació en Intel.ligència Artificial (IIIA). PhD Thesis, 1997.

[6] North, D.C. *Institutions, Institutional Change and Economics Performance*. Cambridge U. P., 1990.

[7] Parunak, H. V. D. and Odell, J. Representing social structures in uml. In *Proceedings of the Agent-Oriented Software Engineering Workshop*. Held at the Agents 2001 Conference, 2001.

[8] Rodríguez-Aguilar, J. A. *On the Design and Construction of Agent-mediated Institutions*. PhD thesis, Autonomous University of Barcelona, 2001.

[9] Rodríguez-Aguilar, J. A., Martín, F. J., Noriega, P., Garcia, P., and Sierra, C. Competitive scenarios for heterogeneous trading agents. In *Proceedings of the Second International Conference on Autonomous Agents (AGENTS'98)*, pages 293–300, 1998.

[10] Rodríguez-Aguilar, J. A., Noriega, P., Sierra, C., and Padget, J. Fm96.5 a java-based electronic auction house. In *Second International Conference on The Practical Application of Intelligent Agents and Multi-Agent Technology(PAAM'97)*, pages 207–224, 1997.

[11] Wooldridge, M., Jennings, N. R., and Kinny, D. A methodology for agent-oriented analysis and design. In *Proceedings of the Third International Conference on Autonomous Agents (AGENTS'99)*, 1999.

Chapter 33

EMBODIED CONVERSATIONAL AGENTS IN E-COMMERCE APPLICATIONS

Helen McBreen
Centre for Communication Interface Research, The University of Edinburgh

Abstract This section discusses an empirical evaluation of 3D embodied conversational agents, in three interactive VRML e-commerce environments: a cinema box-office, a travel agency and a bank. Results showed participants enjoyed speaking to the agents and expressed a desire for agents in the cinema to be informally dressed but those in the bank to be formally dressed. Qualitative results suggested that participants found it difficult to assign a degree of trust to the agents in the banking application.

1. Introduction

The emerging interest in embodied conversational agents (ECA's) coupled with the growing evidence [1, 3, 4, 6, 9, 12] that embodiment can enhance user interface design has fuelled a challenging research agenda and developing embodied agents that behave socially in an interaction has become the principal goal for many interdisciplinary researchers involved with the development of intelligent communicative systems. Virtual Reality Modelling Language (VRML) is an effective tool to describe 3D environments increasing the information density for the user and adding additional layers of perception and meaning to the experience [5]. Inhabiting 3D environments with 3D embodied agents and endowing these agents with conversational capabilities can promote an effective social interaction. Cassell et al [6] have explored the affordances of embodiment and showed that an ECA can improve the interaction and the experience for the user because the agent "enables the use of certain communication protocols in face-to-face conversation which provide for a more rich and robust channel of communication than is afforded by any other medium available today".

Hayes-Roth [7] has proposed that the Internet should be inhabited with smart interactive characters that can engage users with social communication skills as in the real world, enhancing mundane transactions and encouraging a sense of presence for the user, resulting in more effective and efficient interaction. Developing further this proposal, Ball [3] demonstrated that endowing animated agents with personality and emotion creates a sense of social presence, leading to more useful conversational interfaces. The existence of this social presence is important in order to begin to understand the development of the interaction between the agent and the user. It follows from this that understanding the creation and development of social relationships between the agents and the users is a crucial first step to creating socially intelligent embodied conversational agents.

There is little empirical evidence yet available to demonstrate the effectiveness of ECA's, particularly in e-commerce applications and there is a growing need for the establishment of objective and subjective measures of usability. Ostermann [10] developed an architecture designed to support e-commerce "by providing a more friendly, helpful and intuitive user interface compared to a regular browser". Results from experiments using this architecture showed that facial animation was favoured over text only interfaces. These results are encouraging, but it is also necessary to investigate the range of applications that can be significantly enhanced by the presence of an ECA and what are users' attitudes toward their appearance, personality and trustworthiness during the interaction.

The goal of this study is to present empirical evidence in support of the use of the agents within e-commerce domains, in addition to documenting qualitative and quantitative data regarding users' subjective experience of successive interactions with the agents. A detailed discussion of the experimental findings is obviously beyond the scope of this section, however the experimental procedure, key findings and challenge problems are presented.

2. Experimental Research

This experiment assessed two types of 3D male and female embodied agents, appearing as assistants in VRML e-commerce applications (cinema, travel agency and bank). The agents types were a smartly dressed (formal) agent and a casually dressed (informal) agent. In order to evaluate the agents, a real-time experimental platform system, capable of face-to-face conversation between the user and the agent was used.

The first prediction was that participants would believe ECA's have a role to play as assistants. This prediction was made based on the results of previous experiments, where customers passively viewed conversational agents in retail spaces [9] and indicated a desire to actually converse with them. A second prediction was that participants would enjoy speaking to the agents equally in all three applications. This prediction was made based on the fact that the agents were designed to offer the same enhancement in each application, i.e. assisting the user with their tasks. Thirdly, it was hypothesised that the stereotypes created (formal and informal) would be better suited to different application environments. In general assistants in cinema box offices dress casually and those in banks more formally. It was predicted that the situation in the virtual environments would mirror these real life scenarios. Finally, as the verbal and non-verbal behaviour for all the agents was identical it was predicted that attitudes to the agents' functionality, aspects of personality and trustworthiness would be similar within and between the applications.

2.1 Experimental Platform Design

The system architecture is based on a client-server system. Using a speech recogniser, the users speech input is captured on the client PC. A Java-based dialogue manager controls the direction of the dialogue as the user completes a task in each application. The 3D applications (Figure 33.1) were created using VRML97, the international standard file format for describing interactive 3D multimedia on the Internet. The VRML code is stored on the server PC.

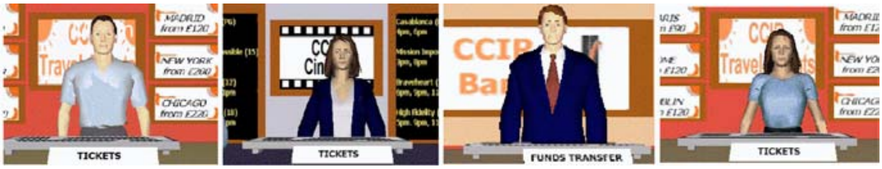


Figure 33.1. Images of ECA's in Applications

The embodied agents were created using MetaCreations Poser 4.0, a character animation software tool. The agents were exported to VRML97 where the code was fitted to the H-Anim specification template [11]. This specification is a standard way of representing humanoids in VRML97. Using this specification it was possible to obtain access to the joints of the agent to create gestures and mouth movements. Four gestures were created for the embodied agents: nodding, waving, shrugging and typing. One male and one female voice recorded the necessary output prompts for the male and female agents respectively. All four agents had the same verbal output.

2.2 Experimental Procedure

Participants (N = 36) were randomly assigned all conditions in a 2 x 2 x 3 repeated measures design: agent gender (male, female), agent type (formal, informal), application (cinema, travel, bank). The presentation of the agents to the participants was randomised within the applications and applications were balanced amongst the participants. Participants were distributed equally according to gender and age group (age 18-35, 36-49, 50+).

Participants were told they would be asked to speak to assistants to complete short tasks in the applications. In all cases the participants were asked to carefully observe the assistant and the application. After the conversation participants completed a 7-point Likert [8] attitude questionnaire relating to the assistant. When participants had seen all four agents in an application they filled out a questionnaire relating to the application. After participants had interacted with all agents in all three applications they completed a questionnaire stating their application preference. A structured interview followed.

2.3 Experimental Findings

2.3.1 E-Commerce Applications. The mean rating scores from the 10-point (low-high) application rating scale show a largely positive response to the applications. No effects for between-subject variables of age and gender were found. A 3 x 1 repeated measures ANOVA taking experimental application as the independent variable showed no significant effects for applications ($F = 0.76, df = 2.0, p = 0.47$). The cinema was rated the highest, followed by the travel agency and thirdly the bank (mean score: cinema = 6.56; travel = 6.46; bank = 6.12). The 7-point Likert questionnaire used to retrieve information about the participants' attitudes toward the applications showed participants felt the applications were convenient and easy to use.

A chi-square test showed the cinema application was significant preferred in comparison to the other applications ($p < 0.05$). In fact, 40% of participants preferred the cinema application, 14% of participants preferred the travel agency and 14% preferred the banking application. A further 8% did not like any of the applications and 25% of the participant sample liked all applications equally.

One participant commented the experience was an improvement because of the feeling of "dealing with someone face to face" and the cinema application "seemed easier to use". In all three applications participants experienced delayed responses from the system as it was processing information and the general thought was that if the delays could be eliminated, the applications would be more successful. The delays seemed to reduce user confidence is the systems, especially where more critical information was being inputted (travel, bank). Participants were also uncertain about security, confidentiality and reliability when completing transactions in the banking application. It was suggested that more visual content in the form of text output would be an improvement. Also, having the opportunity to use the keyboard to enter security numbers may be a beneficial feature.

2.3.2 Embodied Conversational Agents. A series of repeated measures 2 x 2 x 3 ANOVAs taking agent gender, agent type and application as the within-subject independent variables were conducted to analyse participants' attitudes to the questionnaire items relating to the embodied agents as assistants. The questionnaire addressed key issues relating to the agents' personality, trustworthiness and appearance.

All the agents were perceived as being equally friendly and competent. In addition all four agents were perceived as being sociable, cheerful, and agreeable. Participants were asked if the assistants were trustworthy. Although just approaching significance ($F = 2.97, df = 2.0, p < 0.06$), the mean results did show that the assistants in the bank scored less than the assistants in the other applications (mean score: cinema = 5.15; travel = 5.23; bank = 4.93).

Results showed (Figure 33.2) significant preference for the formal agents in the banking application, ($p < 0.01$). Significant results (Figure 33.3) also showed participants felt it would be more appropriate for agents in the cinema application to be dressed informally and agents in the banking application to be dressed formally, ($F = 15.65, df = 2.0, p < 0.01$).

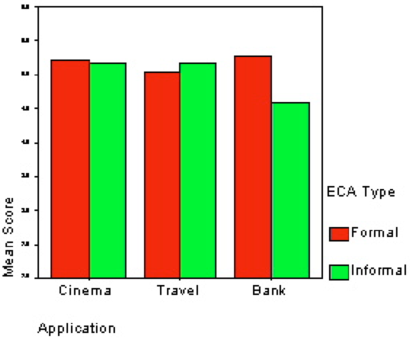


Figure 33.2. Attitude to Appearance

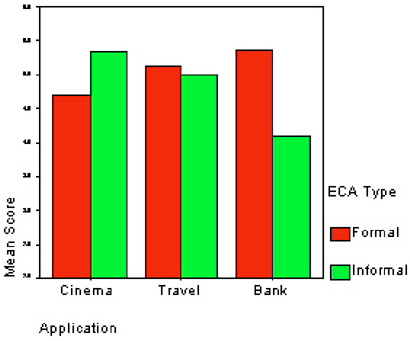


Figure 33.3. Attitude to Appropriateness of Assistants Dress

All participants in the experiment took part in a structured interview. Many comments suggested ways to improve the system. Participants felt that the agents' gesturing was at times "a bit awkward". This highlights one of the challenge problems of creating autonomous animated embodied agents with fluid movements. Research in on-going to address this issue. For instance Badler [2] is using parallel transition networks as a mechanism to create realistic movement for animated agents.

Due to real-time technological restraints, some of the output responses were delayed and participants found these delays off-putting and annoying, giving the impression that the assistant seemed unsure. This highlights another challenging problem within the area of ECA research. With technological improvements this issue may be resolved, improving user confidence with respect to the security, confidentiality and reliability of such systems.

Two thirds of the participants (24/36) thought the assistants enhanced the services and they enjoyed speaking to them. One participant said: “I enjoyed talking to the assistants, I was even polite to them”. Participants felt the assistants should be polite and cheerful, demonstrating competence during the interaction. To do this it was suggested that they should smile and provide appropriate verbal and non-verbal feedback.

3. Discussion

It was hypothesised that participants would respond positively to the embodied agents. The results support this prediction suggesting that 3D ECA's have a role to play as assistants in VRML e-commerce applications. The results supported also a further claim that casually dressed agents are more suitable in virtual cinemas, and formally dressed agents are more suitable in virtual banking applications. It is important to know that ECA's would be welcomed in e-commerce domains especially given the number of commercial websites that are exploring the use ECA's as marketing tools (e.g. Extempo Inc, VirtualFriends).

Participants felt the cinema was more entertaining than the travel agency and banking application. Although ECA's were welcomed in all three retail applications, results suggest it is important to consider carefully the nature of the application task and be aware that ECA's might be more effective in less serious applications, where the consequences of failure are less serious. Nevertheless, the responses to the use of ECA's in these more serious applications may be improved if users' confidence in the system can be increased and the trustworthiness of the agent can be firmly established. Suggested methods to achieve this included better and faster response times from the agents, having the opportunity to enter data using the keyboard and also seeing additional textual feedback on the interface.

All four agents were perceived to be polite, friendly, competent, cheerful, sociable and agreeable; all traits important for assistants in retail and e-commerce spaces. The trustworthiness of the agents was the only aspect where differences between the applications emerged. The qualitative results showed that participants were less likely to trust agents to complete tasks correctly in the banking application. During the interviews, participants stated that they would be more likely to use the applications if the ECA was more convincing that the inputted information was being processed correctly.

4. Conclusions & Future Research

Establishing trust between the agent and the user is of great importance, and on-going research [4] is exploring the construction of a social relationship to assist with establishing trust. Unless users are confident that the agent can understand and process information correctly they may be less likely to trust it, resulting in a less effective interaction. In the study by van Mulken et al [12] results showed personification of interfaces does not appear to be sufficient for raising trustworthiness. If this is the case what other methods could be used for establishing trust in e-commerce applications?

The use of text in the interface could be used to provide feedback to the user about the information the agents have received and processed and may improve user confidence. Allowing the use of keyboard entry in conjunction with speech input, especially when entering security details may also be an improvement. Using the same experimental platform described for this experiment, text-input and text-output will be added to the system in order to further the research aspects of user confidence to ECA's in e-commerce applications. Research suggests the development of ECA's in all domains will be dictated not only by technological advances but also by advances in the understanding and creation of the social interaction between the agent and user, in particular the establishment of trust.

Acknowledgments

Thanks to colleagues at CCIR for helpful comments, in particular Prof. M.A. Jack and Dr. J.C. Foster. Sincere gratitude is also expressed to Dr. J.A. Anderson for developing the dialogue manager software.

References

[1] E. Andre and T. Rist. Personalising the user interface: Projects on life-like characters at DFKI. In Proc. 3rd Workshop on Conversational Characters, 167–170, October 1998.

[2] N. Badler, R. Bindiganavale, J. Allbeck, W. Schuler, L. Zhao, and M. Palmer. Parameterized action representation for virtual human agents. In J. Cassell, et al. (eds.), *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[3] G. Ball and J. Breese. Emotion and personality in a conversational agent. In J. Cassell, et al. (eds.), *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[4] T. Bickmore and J. Cassell. How about this weather? Social dialogue with embodied conversational agents. In Proc. AAAI Fall Symposium: Socially Intelligent Agents, 4–8, November 2000.

[5] M. Bricken. Virtual worlds: No interface to design. Technical Report R-90-2. Washington Technology Center, WA, 1990.

[6] J. Cassell, et al. (eds.). *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[7] B. Hayes-Roth. Characters everywhere. Seminar on People, Computers and Design, March 2001. Stanford University.

[8] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 140:55–62, 1932.

[9] H. McBreen and M. Jack. Empirical evaluation of animated agents in a multi-modal retail application. In *Proc. AAAI Fall Symposium: Socially Intelligent Agents*, 122–126, November 2000.

[10] J. Ostermann and D. Millen. Talking heads and synthetic speech: An architecture for supporting electronic commerce. In Proc. IEEE Int. Conf. On Multimedia and Expo, 2000.

[11] S. Ressler, C. Ballreich, and M. Beitler. Humanoid Animation Working Group, 2001. http://www.h-anim.org.

[12] S. van Mulken, E. Andre, and J. Muller. An empirical study on the trustworthiness of life-like interface agents. In Proc. HCI'99: Communication, Cooperation and Application Design, 152–156, 1999.

Index