

Animating Expressive Characters for Social Interaction

Advances in Consciousness Research (AiCR)

Provides a forum for scholars from different scientific disciplines and fields of knowledge who study consciousness in its multifaceted aspects. Thus the Series includes (but is not limited to) the various areas of cognitive science, including cognitive psychology, brain science, philosophy and linguistics. The orientation of the series is toward developing new interdisciplinary and integrative approaches for the investigation, description and theory of consciousness, as well as the practical consequences of this research for the individual in society. From 1999 the Series consists of two subseries that cover the most important types of contributions to consciousness studies:

Series A: Theory and Method. Contributions to the development of theory and method in the study of consciousness; Series B: Research in Progress.

Experimental, descriptive and clinical research in consciousness.

This book is a contribution to Series B.

Editor

Maxim I. Stamenov

Bulgarian Academy of Sciences

Editorial Board

David J. Chalmers
Australian National University

Gordon G. Globus
University of California at Irvine

Christof Koch
California Institute of Technology

Stephen M. Kosslyn
Harvard University

Stephen L. Macknik
Barrow Neurological Institute, Phoenix, AZ,
USA

George Mandler
University of California at San Diego

Susana Martinez-Conde
Barrow Neurological Institute, Phoenix, AZ,
USA

John R. Searle
University of California at Berkeley

Petra Stoerig
Universität Düsseldorf

Volume 74

Animating Expressive Characters for Social Interaction

Edited by Lola Cañamero and Ruth Aylett

Animating Expressive Characters for Social Interaction

Edited by

Lola Cañamero

Ruth Aylett

Heriot-Watt University

John Benjamins Publishing Company

Amsterdam / Philadelphia



™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Animating expressive characters for social interaction / edited by Lola Cañamero and Ruth Aylett.

p. cm. (Advances in Consciousness Research, ISSN 1381-589X ; v. 74)

Includes bibliographical references and index.

1. Social interaction. 2. Emotions. I. Cañamero, Lola.

HM1111.A55

2008

302.2'22--dc22

2008033085

ISBN 978 90 272 5210 4 (Hb; alk. paper)

© 2008 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

To Fiorella de Rosis, *in memoriam*.

Her sharp mind, kindness, strength, high standards of commitment, as well as her support for interdisciplinary collaboration, new ideas and young researchers were a fundamental driving force in affective computing and other communities. Her memory will continue to be a source of inspiration for all of us.

Table of contents

About the editors	IX
List of contributors	XI
Introduction	XV
<i>Lola Cañamero and Ruth Aylett</i>	
1. Social emotions	1
<i>Paul Dumouchel</i>	
2. Fabricating fictions using social role	21
<i>Lynne Hall and Simon Oram</i>	
3. What's in a robot's smile? The many meanings of positive facial display	37
<i>Marianne LaFrance</i>	
4. Facial expressions in social interactions: Beyond basic emotions	53
<i>Susanne Kaiser and Thomas Wehrle</i>	
5. Expressing emotion through body movement: A component process approach	71
<i>Mark Coulson</i>	
6. Affective bodies for affective interactions	87
<i>Marco Vala, Ana Paiva and Mário Rui Gomes</i>	
7. Animating affective robots for social interaction	103
<i>Lola Cañamero</i>	
8. Dynamic models of multiple emotion activation	123
<i>Valeria Carofoglio, Fiorella de Rosis and Roberto Grassano</i>	
9. Exercises in style for virtual humans	143
<i>Zsófia Ruttkay, Catherine Pelachaud, Isabella Poggi and Han Noot</i>	

10. Expressive characters in anti-bullying education <i>Ruth Aylett, Ana Paiva, Sarah Woods, Lynne Hall and Carsten Zoll</i>	161
11. Psychological and social effects to elderly people by robot-assisted activity <i>Takanori Shibata, Kazuyoshi Wada, Tomoko Saito and Kazuo Tanie</i>	177
12. Designing avatars for social interactions <i>Marc Fabri, David J. Moore and Dave J. Hobbs</i>	195
13. Applying socio-psychological concepts of cognitive consistency to negotiation dialog scenarios with embodied conversational characters <i>Thomas Rist and Markus Schmitt</i>	213
14. Semi-autonomous avatars: A new direction for expressive user embodiment <i>Marco Gillies, Daniel Ballin, Xueni Pan and Neil A. Dodgson</i>	235
15. The Butterfly effect: Dancing with real and virtual expressive characters <i>Lizbeth Goodman, Ken Perlin, Brian Duffy, Katharine A. Brehm, Clilly Castiglia and Joel Kollin</i>	257
16. The robot and the baby <i>John McCarthy</i>	279
Subject index	293

About the editors

Lola Cañamero (<http://homepages.feis.herts.ac.uk/~comqlc>) is Reader in Adaptive Systems at the School of Computer Science of the University of Hertfordshire (UH), United Kingdom, where has been faculty since 2001. She received a BA and MA in Philosophy from the Complutense University of Madrid, and a PhD in Computer Science (1995) from the University of Paris-XI. She worked as a post-doc in the group of Rodney Brooks at the MIT AI-Lab (1995–1996) and in the group of Luc Steels at the VUB AI-Lab (1997), and as senior researcher at the Spanish Scientific Research Council (1998–2000). Since 1995, her research has revolved around affect (motivation and emotion) modeling for autonomous and social agents/robots and adaptive behavior. At UH, she currently leads research on these topics and their intersections with other areas such as developmental and embodied robotics and human-robot interaction, focusing particularly on: embodied architectures based on motivations and emotions for decision-making in autonomous robots; motivation- and emotion-based learning of affordances; artificial evolution of affective systems; the role of affect in imitation; the development of affective bonds in robots and in simulated social groups; and expressive robotic heads for the study of emotion development and social interactions. She has organized over 14 international conferences and workshops in Europe and the USA since 1994, and acted as PC member of over 30 in these areas. She is author or co-author with her students over 80 refereed scientific papers, co-editor of the book *Socially Intelligent Agents: Creating relationships with computers and robots* (Kluwer Academic Publishers, 2002), guest editor (with Paolo Petta) of the special issue of *Cybernetics and Systems: An International Journal* “Grounding Emotions in Adaptive Systems” (2001), and of the special issue of the *International Journal of Humanoid Robotics* “Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids” (2006, with Catherine Pelachaud). She is member of the Editorial Board of the journal *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* (John Benjamins). She has been a full member of the International Society for Research on Emotion (ISRE) since 1999. Between January 2004 and December 2007, she coordinated the area “Emotion in Cognition and Action” of the EU-funded HUMAINE Network of Excellence (<http://emotion-research.net>), and since December 2006 she coordinates the also

EU-funded Advanced Robotics project FEELIX GROWING (<http://www.feelix-growing.org>) on socially situated emotional development.

Ruth Aylett (<http://www.macs.hw.ac.uk/~ruth>) has been a Professor of Computer Science at Heriot-Watt University In Edinburgh since 2004, where she leads the VIS&GE research group (Vision, Interactive Systems and Graphical Environments). This followed an initial job in the then British Computer industry of the late 1970s, and posts in Sheffield University, Sheffield Hallam University, University of Edinburgh and University of Salford.

Her research concerns the overlap of artificial intelligence and real-time interactive graphics, specifically affective agent architectures and interactive narrative. She has developed the idea of emergent narrative as an approach to solving the conflict between user freedom and narrative structure in interactive graphics environments and worked on believable characters able to sustain this approach. She has coordinated successive EU-funded projects since 2002 applying these ideas to a virtual drama system containing intelligent autonomous characters for educating against bullying behaviour in schools, and recently to education in intercultural empathy. Approaches to reconciling the psychology-based cognitive appraisal approach to affect with more neuro-physiological and somatic accounts are a long-term interest.

She is currently a partner in the EU project LIREC – Living with Robots and intEractive Characters – which is investigating how robots and graphical characters can become long-term companions to their human users. This includes work on long-term memory organised around auto-biographical episodes and indexed by emotional state. Other virtual character-based work includes a mobile guide ‘with attitude’ running on a hand-held device.

She has published more than 150 articles as book chapters, journal papers and refereed conference papers and taken part in a large number of international conference programme committees, most recently acting as a joint programme chair of ACII2007 and a senior programme committee member of AAMAS 2008. She was the initiator of the international conference Intelligent Virtual Agents (IVA). She has also published a popular science book *Robots: Bringing Intelligent Machines to Life?*

List of contributors*

Ruth Aylett
Mathematics and Computer Science
Heriot-Watt University
Edinburgh EH14 4AS
United Kingdom
Email: ruth@macs.hw.ac.uk
<http://www.macs.hw.ac.uk/~ruth/>

Daniel Ballin
Global Mobility, BT Global Services
HW N547, P.O. Box 800
London N18 1YB
United Kingdom
Email: daniel.ballin@bt.com

Katharine A. Brehm
SMARTlab Digital Media Institute &
MAGICGamelab
University of East London
4-6 University Way
London E16 2RD
United Kingdom

Lola Cañamero
School of Computer Science
University of Hertfordshire
College Lane
Hatfield, Herts AL10 9AB
United Kingdom
Email: L.Canamero@herts.ac.uk
<http://homepages.feis.herts.ac.uk/~comqlc>

Valeria Carofiglio
Intelligent Interfaces Research Group
Department of Computer Science
University of Bari
Via E. Orabona 4, 70123 Bari
Italy
Email: carofiglio@di.uniba.it
<http://www.di.uniba.it/intint/people/valeria.html>

Clilly Castiglia
Brand Experience Lab
520 Broadway, 5th Floor
New York, NY 10012, USA
<http://clillyc.googlepages.com>

Mark Coulson
School of Health and Social Sciences
Middlesex University
Queensway
Enfield EN3 4SF
United Kingdom
Email: m.coulson@mdx.ac.uk
<http://www.middlesex.ac.uk/hssc/staff/profiles/academic/coulsonm.asp>

Neil Dodgson
University of Cambridge
Computer Laboratory
William Gates Building
15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
Email: nad10@cl.cam.ac.uk
<http://www.cl.cam.ac.uk/~nad10/>

* We regret that Professor Fiorella de Rosis, formerly of University of Bari in Italy, and Professor Kazuo Tanie, formerly of the National Institute of Advanced Science and Technology (AIST) in Japan, died before this book was published.

Brian Duffy
CoE Technology, AGS
805 Av. Dr. Maurice Donat
06254 Mougins Cedex
France
Email: Brian.Duffy@manmachine.org

Paul Dumouchel
Graduate School of Core Ethics
and Frontier Sciences
Ritsumeikan University
56-1 Toji-in Kitamachi
Kita-ku, Kyoto 603-8577
Japan
Email: dumouchp@ce.ritsumei.ac.jp

Marc Fabri
Innovation North
Faculty of Information and Technology
Leeds Metropolitan University
Headingley Campus
Leeds LS6 3QS
United Kingdom
Email: m.fabri@leedsmet.ac.uk
<http://www.lmu.ac.uk/ies/comp/staff/mfabri/index.htm#home>

Marco Gillies
Department of Computing
Goldsmiths, University of London
New Cross
London SE14 GNW
United Kingdom
Email: m.gillies@gold.ac.uk

Mário Rui Gomes
IST/UTL
ST Taguspark
Av. Prof. Dr. Cavaco Silva
2744-016 Porto Salvo
Portugal
Email: mario.gomes@tagus.ist.utl.pt

Roberto Grassano
Intelligent Interfaces Research Group
Department of Computer Science
University of Bari

Via E. Orabona 4, 70123 Bari
Italy
Email: grassano@di.uniba.it

Lizbeth Goodman
Creative Technology Innovation
SMARTlab Digital Media Institute &
MAGICGamelab
University of East London
4-6 University Way
London E16 2RD
United Kingdom
<http://www.smartlab.uk.com>

Lynne Hall
School of Computing & Technology
University of Sunderland
Sunderland, SR6 0DD
United Kingdom
Email: lynne.hall@sunderland.ac.uk

David Hobbs
University of Bradford
School of Informatics
Department of EIMC
Bradford, West Yorkshire BD7 1DP
United Kingdom
Email: d.hobbs@bradford.ac.uk
<http://www.inf.brad.ac.uk/~dhobbs/>

Susanne Kaiser
Department of Psychology
University of Geneva
40, Boulevard du Pont-d'Arve
CH-1205 Geneva
Switzerland
Email: Susanne.Kaiser@pse.unige.ch
<http://www.unige.ch/fapse/emotion/members/kaiser/kaiser.htm>

Joel Kollin
DXARTS
Box 353414
University of Washington
Seattle, WA 98195-3414
USA

Marianne LaFrance
Department of Psychology
Yale University
P.O. Box 208205
New Haven, CT 06520-8205
USA
Email: marianne.lafrance@yale.edu
<http://www.yale.edu/psychology/FacInfo/LaFrance.html>

John McCarthy
Computer Science Department
Stanford University
Stanford, California 94305-9020
USA
Email: jmc@cs.stanford.edu
<http://www-formal.stanford.edu/jmc/personal.html>

David Moore
Innovation North
Faculty of Information and Technology
Leeds Metropolitan University
Headingley Campus
Leeds LS6 3QS
United Kingdom
Email: d.moore@leedsmet.ac.uk
<http://www.leedsmet.ac.uk/inn/DavidMoore.htm>

Han Noot
Centrum voor Wiskunde
en Informatica (CWI)
Kruislaan 413, 1098 SJ,
P.O. Box 94079, 1090 GB Amsterdam
The Netherlands
Email: Han.Noot@cw.nl

Simon Oram
Lead developer
ElectroSoup
United Kingdom
<http://www.electrosoup.co.uk/>

Ana Paiva
INESC-ID / Instituto Superior Técnico
Sala 2-N7.23
Tagus Park
Av. Prof. Dr. Cavaco Silva
2780-990 Porto Salvo
Portugal
Email: ana.paiva@inesc-id.pt
<http://gaips.inesc-id.pt/gaips/indiv.php?id=1>

Xueni Pan
Department of Computer Science
University College London
Malet Place
London WC1E 6BT
United Kingdom
Email: s.pan@cs.ucl.ac.uk
<http://www.cs.ucl.ac.uk/staff/S.Pan/>

Catherine Pelachaud
IUT de Montreuil, Université de Paris 8
140 rue de la Nouvelle France
93100 Montreuil
France
Email: pelachaud@iut.univ-paris8.fr
<http://www.iut.univ-paris8.fr/~pelachaud/>

Currently on leave at:
TELECOM, ParisTech/TSI
46, rue Barrault
75634 Paris Cedex 43
France
Email: catherine.pelachaud@telecom-paristech.fr

Ken Perlin
New York University
Media Research Laboratory
Dept. of Computer Science
719 Broadway, Room 1202
New York, NY 10003, USA
Email: perlin@mrl.nyu.edu

Isabella Poggi
Università Roma Tre
Dipartimento di Scienze dell'Educazione
Via del Castro Pretorio, 20
00185 Roma
Italy
Email: poggi@uniroma3.it
<http://host.uniroma3.it/docenti/poggi/>

Thomas Rist
University of Applied Sciences Augsburg
Baumgartnerstr. 16
86161 Augsburg
Germany
Email: tr@rz.fh-augsburg.de

Zsofia Ruttkay
Human Media Interaction (HMI)
Dept. of Electrical Engineering,
Mathematics and Computer Science
University of Twente
Zilverling (building 11) room 2033
The Netherlands
Email: zsofi@cs.utwente.nl
<http://wwwhome.cs.utwente.nl/~zsofi/>

Tomoko Saito
1-1-1 Umezono
Tsukuba
Ibaraki 305-8568
Japan
Email: tomo-saito@aist.go.jp

Markus Schmitt
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken
Germany

Takanori Shibata
Bio-Robotics Division, Robotics Department
Mechanical Engineering Laboratory (MEL)
Ministry of Economy, Trade and Industry
(METI)
1-2 Namiki, Tsukuba 305
Japan
Email: shibata@mel.go.jp

<http://www.aist.go.jp/MEL/soshiki/robot/biorobo/shibata/shibata.html>

Marco Vala
INESC-ID / Instituto Superior Técnico
Sala 2-N7.23
Tagus Park
Av. Prof. Dr. Cavaco Silva
2780-990 Porto Salvo
Portugal
Email: marco.vala@inesc-id.pt

Kazuyoshi Wada
Faculty of System Design
Tokyo Metropolitan University
6-6 Asahibaoka, Hino
Tokyo 191-0065
Japan
Email: k_wada@sd.tmu.ac.jp

Thomas Wehrle
Universität Zürich
Psychologisches Institut
Psychologische Methodenlehre
Binzmühlestrasse 14/27
CH-8050 Zürich
Switzerland
Email: t.wehrle@psychologie.unizh.ch
<http://tecfa.unige.ch/perso/wehrle/>

Sarah Woods
School of Psychology
University of Hertfordshire
College Lane, Hatfield
Hertfordshire, AL10 9AB
United Kingdom
Email: s.n.woods@herts.ac.uk

Carsten Zoll
Allgemeine Psychologie/General Psychology
Otto-Friedrich-Universität
Kapuzinerstraße 16
96045 Bamberg
Germany
Email: carsten.zoll@uni-bamberg.de

Introduction

Lola Cañamero and Ruth Aylett

1. Motivation and background

The ability to express and recognize emotions is a fundamental aspect of social interaction. The importance of endowing artifacts (synthetic characters or robots) with these capabilities is nowadays widely acknowledged in different research areas such as affective computing, socially intelligent agents, computer animation, or virtual environments, and thus the interest in emotions and their expression spans very different disciplines. Expressing and recognizing emotions is studied in the social agent community because it is known to be fundamental to establishing a social relationship; it is studied in computer animation because it is also a basic requirement for the believability of animated characters and for human engagement in the narratives in which they are involved. It has also been studied for many hundreds of years in expressive arts and, for most of the 20thC and into the 21stC, in the psychology of emotion. With the growth of synthetic characters in virtual environments and on the web, as well as of the introduction of domestic or entertainment robots, this topic is also receiving an increased importance in various areas of computer science, artificial intelligence and related disciplines. However, these different communities study emotional expression and interactions in different ways and often do not interact with each other due, amongst other reasons, to the lack of appropriate platforms.

Even within computing, there is a gulf between graphics and animation researchers who are concerned about exterior expressiveness, and workers in AI and cognitive science who build computer models of the internals of such artifacts. Meanwhile communities entirely outside of computing have ideas known little or not at all within it but of major potential use, as witness the recent use of Laban analysis from dance choreography in behavioral animation. Researchers in all these areas are however confronted with the problem of how to make the emotional displays of artifacts and characters believable and acceptable to humans. This can involve not only generating appropriate expressions and behavioral displays – explored in animated film for many years – but also endowing artifacts

with underlying models of personality and emotions that support the coherence and autonomy of their emotional displays and interactions.

Our motivation behind this book and the AISB'02 symposium *Animating Expressive Characters for Social Interactions*,¹ from which the idea of this book arises, was to take a step towards bridging this gap, in two respects:

1. Multi-disciplinarity: Bringing together work from different disciplines (including psychology, the arts, computer graphics and animation, socially intelligent agents, synthetic characters, robotics, virtual reality, etc.) to reflect on this common problem from different perspectives and to gain new insights from this multi-disciplinary feedback.
2. "Animation" as unifying focus. Although different events and publications have explored isolated aspects of emotions and their expression in artificial agents (e.g., emotion-based architectures, models of personality, believability, interfaces, etc.) and of animation in the sense this term has in the graphics community, to our knowledge no single event or publication has brought together the different aspects, models and techniques involved in designing and animating expressive characters for social interactions, from internal mechanisms to external displays. Our book concerns "animation" not only from a graphical perspective, but more generally in the human sense: making characters "life-like", externally but also "internally" – giving them an "anima", so that they appear as life-like entities and social partners to humans.

This book presents a multi-disciplinary collection of articles on various aspects (models and techniques) involved in animating (in the broad sense of the term mentioned above) synthetic and robotic expressive characters for social interactions. Its appropriateness for this particular series, *Advances in Consciousness Research*, lies in the interpersonal nature of human cognition, and the way in which relevant features and mechanisms used to "animate" artifacts externally or internally can be used to make them appear as life-like entities and social partners to humans. Expressive behavior is a basic element in the functioning of the 'Theory of Mind' that allows us to infer motivations, intentions and goals in other humans, and thus perhaps by extension in graphical and robotic agents and characters.

The intended readership is of a multi-disciplinary nature. This book does not require the reader to possess any specialist knowledge; it is suitable for any student or researcher with a general background in any of the fields informing the overall topic. It can also be used as a reference text / background reading for uni-

1. Information about the AISB'02 Symposium *Animating Expressive Characters for Social Interactions* is available at <http://homepages.feis.herts.ac.uk/~comqlc/aecsi02>

versity courses in a number of topics, such as affective computing, animated and virtual agents, embodied artificial intelligence, animation techniques, etc.

2. Structure of the book and overview of the chapters

The book covers a wide variety of aspects (models and techniques) involved in “animating” (synthetic and robotic) expressive characters for social interactions. Although individual chapters span over various topics, the book can be organized around the following themes, each of them including contributions from different disciplines:

- The social nature of affective interactions: Chapters 1 and 2.
- Expression of emotions: Through the face (Chapters 3 and 4) and through the body (Chapters 5 and 6).
- Internal mechanisms for emotional expression and interaction: Chapters 7, 8 and 9.
- Expressive characters and robots as social partners: Social effects of affective artifacts: Chapters 10 and 11.
- Avatars – Embodying the human user in animated characters: Chapters 12, 13 and 14.
- Emotional interaction in fiction and drama: Chapters 15 and 16.

2.1 Social nature of affective interactions

In Chapter 1, “Social Emotions”, Paul Dumouchel discusses an evolutionary basis for social emotions as a means of regulating interaction within groups of the same species, and assesses Darwin’s seminal work in a modern context. He considers two different views of emotional expressiveness – one that sees it as a way of signaling an internal state, and a second that rather sees it as a definite action in its own right with communicative intent.

In Chapter 2, “Fabricating Fictions Using Social Role”, Lynne Hall and Simon Oram emphasize the often disregarded fact that, like humans, embodied agents work within a culture and carry specific roles in particular situations. They argue that current technology is still far from producing cultural complexity convincingly, and their goal is to create agents that are aware of their social situation, and therefore capable of emitting signals that contextualize them in the world. To this end they have developed PACEO, a personal assistant that organizes meetings in a workplace context, as a tool to explore social interaction between the agent and human users. To build the agent, the approach they take is a pragmatic one, where

believability and user engagement are more important than the “intelligence” of the agent. To construct the agent’s culture, they adopt a Foucauldian perspective to build a narrative of the social space of the agent, particularly concerning the roles that the notions of discourse, bio-power and workplace play in the narrative construction of culture.

2.2 Expression of emotions through the face and the body

Chapter 3, “What’s in a Robot’s Smile? The Many Meanings of Positive Facial Display” by Marianne LaFrance offers a valuable corrective to simplistic accounts of the relationship between facial expression and affective state by discussing the example of smiling. LaFrance shows that smiles appear in a variety of forms in order to express a variety of emotions, with only one, the Duchenne smile, unambiguously associated with happiness. Smiles (and other emotion-related facial displays) are not necessarily indicators of internal states, but often act as social messages, e.g. to show others our disposition towards an interaction episode. From the perspective of smiles as volitional social messages, the problem of distinguishing “true” from “fake” smiles holds less relevance than it has traditionally been given. Instead, understanding the meaning of a smile requires understanding the social context in which that smile occurs, and this chapter examines some of the social dimensions that are related to different types of smiles.

In Chapter 4, “Facial Expressions in Social Interactions: Beyond Basic Emotions”, Susanne Kaiser and Thomas Wehrle illustrate how an appraisal-based approach to understanding the relation between emotion and facial expression can be instrumental to multidisciplinary research that brings together emotion theory and computational modeling, encompassing aspects of facial expression synthesis, automatic expression recognition, and artificial emotions. Going beyond the study of emotions in classical experimental settings, these authors use human-computer interaction, in particular interactive computer games, to study the dynamics of ongoing cognitive and emotional episodes in what they call *emotional problem solving*. In addition to a theoretical model based on a component process approach, Kaiser and Wehrle propose a set of computer tools to perform both, analysis and synthesis of facial emotional expressions using situated coding procedures, and to test predictions arising from the model.

Chapter 5, “Expressing Emotion Through Body Movement: A Component Process Approach”, by Marc Coulson, also takes an appraisal-based, componential approach to the study of emotional expression closely related to that of Kaiser and Wehrle, but this time applied to the study of emotional expression through the body. Contrary to facial emotional expression, systematic studies of bodily

emotional expression are very rare; Coulson identifies a number of reasons for this, such as the greater weight of other bodily functions such as locomotion and manipulation, the higher individual and cultural variation in expression and recognition of emotions through the body, and the higher complexity of the body in terms of degrees of freedom, movements, and postures. The use of a computer simulation of a human body allows Coulson to model a number of functionally significant postures resulting from the outcomes of Stimulus Evaluation Checks, providing a starting point to test not only the quality of the modeling but also of the component process model.

Chapter 6, “Affective Bodies for Affective Interactions” by Marco Vala, Ana Paiva, and Mário Rui Gomes examines the issue of expressive behavior in 3D graphical characters in motion. The authors explore a computer graphics approach in which the posture and stance required for an action can be modified by the affective state of the character to produce expressive behavior in real time. This approach combines neutral animations, stance composition and changes in the speed and amplitude of the animation to create different affective scripts, and can be used in different characters from both humanoid and non-humanoid skeletons. Examples of such scripts are provided in the context of the computer game *FantasyA*, a magic duel in which characters fight each other using spells. Unlike other computer games, the user does not decide which magic spell to use but influences the emotional state of the character, and this chooses spells autonomously. To be able to play, the user must thus recognize the emotional states of the characters.

2.3 Internal mechanisms for emotional expression and interaction

Chapter 7, “Animating Affective Robots for Social Interaction” by Lola Cañamero, considers what artificial emotions can contribute to social interactions between robots and humans. Cañamero argues for a deep, rather than a shallow, emotion system that is grounded into the internal architecture of social robots. She examines both individual and social grounding, including the issues surrounding embodiment, the autobiographic self, and the theory of mind.

Chapter 8, “Dynamic Models of Multiple Emotion Activation” by Valeria Carofiglio, Fiorella de Rosi and Roberto Grassano, applies Dynamic Belief Networks to the problem of an agent “feeling” multiple emotions simultaneously. Such formalism has been chosen to permit the representation of the two main mechanisms by which several emotions that are simultaneously active might interact – they might coexist and mix, or they might switch between each other in rapid succession. The starting hypothesis of this chapter is that emotions are activated by the belief that

an important goal may be achieved or threatened, and its focal point is the change in belief about the achievement (or threatening) of the goals of an agent over time. Using the taxonomy of Ortony, Clore and Collins (1988), they consider which emotions might be mixed together and how emotion elicitation can be linked to appraisal of the state of plans in memory rather than directly to events. To build, test and refine this model, the authors have developed a tool, Mind-Testbed, that they have applied to different domains of varying complexity.

In Chapter 9, “Exercises in Style for Virtual Humans,” Zsófia Ruttkay, Catherine Pelachaud, Isabella Poggi, and Han Noot consider the importance of the overall style – defined as “the stable tendency of choosing nonverbal signals and their performance characteristics in communicating” – of an embodied conversational agent (ECA) to achieve high-quality interaction with a user. Style can be manifested in the usage of verbal and nonverbal signals involving face, head and hands to express a communicative function. In human-to-human communication and interaction, style depends and is affected by multiple factors related to both permanent and contingent goals, social and cultural setting, cognitive capability, emotional state, personality, etc. Drawing from these characteristics to achieve suitable style generation and expression in ECAs, the authors discuss issues relating to the use of a gesture library, including representation and the impact on the design of agent markup languages. They propose a markup language, GESTYLE, to provide a link between the agent’s body and mind. It takes into account style and connects to other tools developed by the authors, such as a reasoning-based system that covers some aspects of style.

2.4 Expressive characters and robots as social partners: Social effects of affective artifacts

In Chapter 10, “Expressive Characters in Anti-Bullying Education,” Ruth Aylett, Ana Paiva, Sarah Woods, Lynne Hall and Carsten Zoll discuss the role of expressive behavior in synthetic characters used as virtual actors. The application FearNot!, involving virtual dramas for education against bullying, has used drama rather than naturalism as its guiding principle to achieve empathy with the synthetic characters, and the chapter considers the impact this has had both on the expressive behavior of the characters and the internal architecture producing it. In anti-bullying education, knowledge of the phenomenon and its negative consequences is not sufficient to meet the pedagogical objectives, since attitudes and emotions are at least as important in order to produce the behavior. For this reason, techniques such as small-group discussion, role-play and dramatic performance by Theatre-in-Education are highly relevant. Results of evaluation in

different countries support the hypothesis that consistency along the main dimensions of empathy is more important than the degree of naturalistic fidelity of expressive behavior.

Chapter 11, “Psychological and Social Effects to Elderly People by Robot-Assisted Activity”, by Takanori Shibata, Kazuyoshi Wada, Tomoko Saito, and Kazuo Tanie, explores the use of a seal-like robot as a tool to reduce anxiety and stress in elderly people in institutions, drawing on results showing that real-world pets do have such a positive effect. The beneficial effects of animal-assisted therapy seem to encompass three main aspects: psychological (effects such as relaxation and motivation), physiological (e.g., improvement of vital signs), and social (e.g., stimulation of communication). The use of animal-like robots instead of real animals opens up the possibility of using this type of therapy in contexts where the presence of animals might be difficult or dangerous. The bulk of the chapter presents evaluation results of empirical studies conducted in institutions over a period of 4 weeks, using both a scripted and a more responsive version of the seal robot, and discusses their effects on the elderly people interacting with them.

2.5 Avatars: Embodying the human user in animated characters

In Chapter 12, “Designing Avatars for Social Interactions”, Marc Fabri, David J. Moore and Dave J. Hobbs emphasize the importance of introducing emotional expression in avatars, since collaborative virtual environments (CVE) technology is a potentially very powerful means of facilitating communication between people working at a distance. These authors investigate how to better represent emotional displays to transmit emotions through the medium of CVE in an effective way, since virtual environments are often poor in terms of the emotional cues that they convey. The approach adopted in this work focuses on computer modeling and empirical evaluation of facial emotional expression of a set of emotion categories, and uses a small set of features rather than a realistic simulation of real-life physiology. They argue that it is not necessary, and may be counter-productive, to assume that a “good” avatar has to be a realistic representation of human physiognomy.

Chapter 13, “Applying Socio-Psychological Concepts of Cognitive Consistency to Negotiation Dialog Scenarios with Embodied Conversational Characters”, by Thomas Rist and Markus Schmitt, investigates how social interaction concepts of cognitive balance, dissonance and congruency can be applied to scenarios in which synthetic characters negotiate on behalf of human users. An application – Avatar Arena – is used as a testbed for exploring these ideas with negotiation dialogues in the context of collaborative meeting arrangements as a sample domain.

Working towards a model that captures some of the group dynamics found in human-human negotiation dialogues, the authors show how dialogues become increasingly richer as a result of incremental extensions of the mind models of the characters. Three versions of such model are presented in this chapter: in a first version the characters only possess knowledge about the domain, in a second version they incorporate attitudes towards some domain concepts (e.g., dates, activities), and in the third version they also hold attitudes towards other characters and knowledge of their liking relationships.

Chapter 14, “Semi-Autonomous Avatars: A New Direction for Expressive User Embodiment” by Marco Gillies, Daniel Ballin and Neil A. Dodgson examines how avatars – the graphical representation of the user in a graphical environment – can be given some amount of low-level autonomous behavior in order to remove some of the control burden from the user. The authors thus bring together ideas from Artificial Intelligence – in particular embodied autonomous agents and action selection – and the more human-interaction-oriented aspects of avatars and virtual characters. After an extensive review of state-of-the-art work in virtual and semi-autonomous characters, and convincing arguments about the interest of semi-autonomous (as opposed to fully human-controlled and fully autonomous) avatars, eye gaze is taken as an example of such a behavior and some examples of how it can be generated by a developed framework for semi-autonomous avatars is presented.

2.6 Emotional interaction in fiction and drama

Chapter 15, “The Butterfly Effect: Dancing with Real and Virtual Expressive Characters” by Lizbeth Goodman, Ken Perlin and Brian Duffy, discusses the use of expressive behavior in robots and graphical characters in conjunction with human dancers as part of an aesthetic dance experience. It describes in particular work carried out to support dancers who are restricted in the use of their own bodies, due to physical impairment.

Chapter 16, “The Robot and the Baby” by John McCarthy, offers a novel – indeed fictional – conclusion to the book. A founding father of the field of artificial intelligence, McCarthy argues against the incorporation of emotional systems into artifacts such as robots, and in this story dramatizes some of his concerns.

Acknowledgements

We are grateful to the participants of the AISB 2002 symposium *Animating Expressive Characters for Social Interactions* for initial discussions that contributed to shape this book, and to all the contributors for their enthusiastic support to this project and their patience during the editorial process. The EU-funded Network of Excellence HUMAINE (FP6-IST-2002-507422) later provided an excellent multi-disciplinary framework for further discussions and exchange of ideas regarding various issues involved in modeling and expressing human and artificial emotions. Last, but not least, we would like to thank Maxim Stamenov for inviting us to edit this book and for his continuing encouragement, and to Bertie Kaal, Patricia Leplae and Hanneke Bruintjes of John Benjamins Publishing Company for their support and patience.

CHAPTER 1

Social emotions*

Paul Dumouchel

Even the thought, ... that others think that we have made an unkind or stupid remark, is amply sufficient to cause a blush, although we know all the time that we have been completely misunderstood. (Darwin 1872:332)

1. Introduction

In what sense are emotions social? Hayek (1988) suggested banishing the term 'social' from the vocabulary of social sciences. He felt that usage has given this word so many different meanings that it no longer constituted a useful tool for communication. Hayek listed 160 different uses of the word 'social' in order to convince us of its terrible polysemy. He also suggested that each time 'social' is associated with a noun, such as 'justice', 'action' or 'geography', a new concept of 'social' is created. Yet 'social' is, as Hayek himself noted, an adjective and if we were to believe that adjectives acquire new meanings each time they are associated with nouns, then we would probably have to give up using them. I think that just as 'a blue door' and 'a blue house' are different phrases that nonetheless use the same concept of 'blue', 'social justice' and 'social action' are two ideas that use the same concept of 'social'. This does not mean that the word 'social' has only one meaning; the term certainly has multiple meanings but the range is far less broad than Hayek feared. Essentially, the various meanings of the term 'social' are in a way like planets. They are not simply wandering stars; on the contrary, they gravitate in an orderly manner around a sun. That sun is the basic meaning of the term 'social' and this meaning is, I believe, related to the specific forms of existence of emotions.

Emotions are social in that they are not the means but the *state* or the *being* of humans living together. The fact that we have an affective life is not a cause of

* This chapter is a revised version of 'Emotions sociales', Chapter 3 of (Dumouchel 1999). Translated from the French by Mary E. Baker.

but *is* the fact that, as beings, we are not completely independent of each other. Emotions have a relationship with our sociability that is not one of causality, but of *identity*, just as in philosophy of mind, for example, it is said that there is identity between a mental state and a physical state of the brain. This is why the way in which emotions are social is primary. It is not derived. Our affective life makes us social animals. It constitutes the social for us. Thus, I will look to emotions to define the meaning of the term 'social'. The adjective 'social' does not take on a special meaning when it is applied to emotions; on the contrary, all the other senses of the word are, or should be, based on that meaning. To say that emotions are social is to say what *the social* is. This is why the social emotions thesis is not relativistic. It does not claim that emotions are constructed socially but that affective life constitutes what is social for us. It targets what is universal in the social: the fact of social organization or the social bond.

This is to say that the social emotions thesis is also an ontological thesis, ontological in a sense that is closer to the way analytical philosophy uses this term than to the way it is understood in the tradition inspired by Hegel or Heidegger, for example. The history of Being, its presence or absence are not in question here. The assertion that emotions are social has to do with the form of existence of emotions, with what they are, with the way that they exist and with what affective life reveals about the kind of beings that we are. A stone, or grass in a field, in other words, any material object, exists in a different way than an event, such as the assassination of Caesar or the sudden arrival of my friend Lukas, and these also exist in a way that is different from a physical disposition, such as fragility, or a psychological attribute, such as suggestibility. Affective life is made up of salient points in a process of co-ordination between people, and these points include things that resemble actions and the results of actions. What we commonly call 'emotions' encompass both of these things and, therefore, the term 'emotion' cannot be considered to refer to a set of homogenous objects. These salient points exist in a manner that is different again from all the entities mentioned above. In particular, as I have said, emotions, seen as salient points in affective life, do not exist as intrinsic individual characteristics, such as height or eye colour. Rather, they resemble being *Grandfather* or *Henri's second cousin*. They are relational properties, i.e., properties that an individual considered in isolation cannot possess and that it makes no sense to predicate of an individual taken alone out of context. For example, a biological characteristic, such as an adaptation, is a relational property. The fact that an organism has an adaptation¹ does not depend

1. It is important not to confuse *being adapted* with *being an adaptation*. *Being an adaptation* is also a relational property, but it is a historical property that does not depend on inter-relations that take place in the present; it depends on the way a characteristic has become established in

on its intrinsic properties alone, e.g., on its physical properties, but also on the relation that it has (or that some of its characteristics have) with its environment (more specifically, with certain aspects of its environment). Nonetheless, and this is fundamental, *having an adaptation* is not a property of the organism and of its environment, but of the organism alone, just as *being taller than someone* is not a property of the taller individual and the shorter individual, but a property of the taller person only, even though it is a comparative property and relational. Having an adaptation is a property of an organism in an environment, just as being taller than is a property of an object in relation to others. The same applies to emotions. Being angry is not an intrinsic physical or psychological property of a subject, but the property of an individual in a certain context, i.e., a relational property. The *social emotions thesis* says that this context is essentially that of the relations agents have with each other. A more precise way of stating this would be 'the relations that they *undertake* with each other'; in other words, emotions do not depend on all the relations among agents but very precisely on those to which they commit themselves, in contrast with the many abstract relations that they can have, such as spatial relations.

Feeling an emotion is however a property of an individual alone. The fact that emotions are relational characteristics does not entail that my emotions are not my own and the fact that they are social does not entail that they are the properties of a group.

I will try to do two things in the rest of this chapter: first, to show that the only area in which emotions could appear as an adaptation is that of strategic intraspecific relations and, second, to show that, as a process of strategic intraspecific co-ordination, emotions have the characteristic of constituting the agents that they co-ordinate. The first assertion can be seen as an evolutionary hypothesis, the second as a developmental hypothesis. In the conclusion, I will distinguish the social emotions thesis from classical theories of moral sentiments that, in their own way, also make emotions the foundation of the social bond, as well as from recent sociological theories of the social construction of emotions.

a population. *Grosso modo*, a characteristic is an adaptation if it has become established in a population because of the advantage it gave to the individuals who shared that characteristic. However, since environments change, it could very well be that an animal's adaptation is no longer adapted to its present situation.

2. Emotions as adaptations in strategic intra-specific relations

Quite frequently, we are treated to the following anecdote² about the lives of certain animals in captivity. It is said that if lions and tigers are put in the same cage, the lions progressively eliminate all the tigers even though individual tigers are much stronger and more powerful than individual lions. The reason for this is that when a lion and a tiger come into conflict, all the lions in the cage come to the aid of their fellow, whereas the other tigers yawn and scratch themselves as the fight goes on. A tiger reacts only when directly threatened. Lions, on the contrary, act as if all the lions in the cage were targeted by what affects any one of them. Lions help each other and show solidarity, which is, one might say, what makes them superior.

This is true, but mutual assistance and solidarity can be misleading when we try to understand what makes lions special in this case. This can be shown simply by asking whether the consequences of a conflict between two lions would be different from those of a conflict between two tigers. The former always threatens to spread throughout the whole community of lions and, if that happens, *adieu* mutual assistance; the conflict can split the group and destroy solidarity. The latter remains local; it is stopped by the supreme indifference of the other tigers. There is no community through which it could spread. The reason why lions help each other is also the reason why they sometimes kill each other. This is what Kant called 'unsociable sociableness' (1784). Lions are social animals whereas the tiger is solitary.

It should not be concluded from this that tigers lack *certain* emotions with which lions are amply provided. On the contrary, I think that we should simply say that tigers lack emotions. Why? Is it not true that tigers roar with rage, fear fire and, if we are to believe Kipling, hate man? Should we believe that they experience no physiological change when tracking and capturing their prey or when fleeing hunters? I am sure that all these situations result in changes in the physical states of tigers and, since tigers have the anatomy of a higher mammal, these changes must resemble what sets off fight or flight in us: a discharge of adrenaline, accelerated heartbeat and increased blood pressure. The problem is not whether these physical states in tigers are followed by or correlated with certain mental states. The reason tigers lack emotions is not that they have an impoverished interior life. I believe the difficulty comes from the fact that, in this case, the physical states that often accompany certain points in emotional life do not occur in an environment that authorizes us to call them emotions. It is not those physical states or

2. It does not matter whether this anecdote is true or false. What is interesting is that it highlights two different types of behaviour.

events that define an emotion, for only the context determines whether or not they are part of a process of affective coordination.

The case resembles that of an animal's ecological fitness when placed in a new environment. The extremely heavy fur of polar bears is certainly well adapted to their natural environment. However, in the Vancouver Zoo it condemns the poor animals to almost complete immobility, which would probably be fatal to them if their essential needs were not taken care of by a human institution. Asserting that the fur of polar bears is just as adapted to the climate of Vancouver and as it is to that of the North Pole because it remains just as long and warm in both places would demonstrate complete failure to understand the concept of fitness. The confusion in this case is between an adaptation, i.e. the polar bear's coat, which is a historical concept, and an animal's ecological fitness, which is a relational concept relative to the animal's present environment. I think that those who base the existence of emotions on the presence of certain physical changes or on certain mental states are making exactly the same sort of category mistake. Just as the ecological fitness of an organism changes in relation to the environment in which it is found, despite no change in the organism's physical characteristics, certain forms of behaviour are or are not points in affective life, depending on the context in which they take place.

My thesis is that intraspecific co-ordination is the fundamental context in which what we call *the expression of emotions* can be considered an adaptation and serves a function that defines it as an emotion. This is the only context in which these physical occurrences can be considered events in affective life.

3. Emotions and their expression

3.1 Darwin on the expression of emotions

It is well known that Darwin (1872) published an important book on the expression of emotions in humans and animals. One of the surprises contained in this book is how little space the author gives to natural selection in his argumentation. At the beginning of the book, Darwin states three principles that he believes regulate and make comprehensible the expression of emotions. Only the first principle – the principle of habit – leaves a little space for natural selection. Darwin says that certain physical movements can be useful in the presence of certain mental states, for example, when a dog curls back its lips and flattens its ears in a fight. Showing its teeth enables it to bite more effectively and flattening its ears reduces the probability that they will be injured. The principle of habit says that the regular association of the mental state related to fighting and these voluntary

physical movements will have the result that when the mental state occurs, 'however feebly', the animal will have a spontaneous tendency to perform the bodily movements associated with that state,³ even if they are not at all useful in the case at hand (Darwin 1872, p. 28). In other words, the slightest beginnings of aggressiveness in a dog will incite it to curl back its lips and flatten its ears, i.e., the animal will be led to express its emotion. This amounts to saying that while utility may have partially presided over the determination of the bodily movements that constitute the expression of emotions and while natural selection could have contributed to the choice of these movements, the expression of emotions itself has no utility and persists despite this absence of usefulness. This is why the role that the principle of habit gives to natural selection remains very limited.

The two other principles, i.e., the principle of antithesis and the principle of direct action of the nervous system, are based neither directly nor indirectly on the possible advantage to the animal of expressing emotions. The principle of antithesis asserts that when, in accordance with the principle of habit, some actions have been associated with a specific mental state, there will be a tendency for exactly the contrary mental state to cause actions in direct opposition⁴ to those associated with the initial mental state, even if those actions are of no utility for the animal. The third principle, that of direct action of the nervous system, simply says that the expression of emotions also includes movements that necessarily result from the way our bodies are made, independent of will or habit (Darwin 1872, pp. 28–29 and 347–348). If an emotion leads, for example, to an increase in the rate of respiration, which consequently causes the nostrils to dilate, this dilation should, according to Darwin, be considered part of the expression of the emotion.

3. The principle of habit says two other important things. First, it asserts that all the movements that form the expression of an emotion today were originally voluntary and that habit made them involuntary; see the remarkable little work by Gauchet (1992) concerning the theoretical approach that makes the voluntary the origin of the involuntary or reflex. Second, the principle of habit asserts that all efforts to suppress the spontaneous habitual expression of an emotion result in movements that are sometimes extremely small but are always expressive themselves.

4. The major problem that this principle faces, and which renders it perfectly empty, is that there is no criterion independent of the expression of emotion that allows one to determine which action is *contrary* to another. Which action is contrary to pricking up one's ears? Bending them? Lowering them? Flattening them? Holding one straight up and the other on an angle? We have no independent criterion that we could use to answer these questions. Indeed, the question is meaningless. There is no contradiction between actions; at most there are obstacles. Given this it is likely that we call an action contrary when we believe it expresses the opposite emotion.

The difficulty for Darwin is that of explaining the purpose of the *expression* of emotions. If the struggle for survival is fundamentally a fight between organisms rather than between groups,⁵ it is hard to see what could be the purpose of the expression of emotions in general. In other words, it is relatively easy to imagine the advantage that certain specific cases of the expression of emotions could procure, such as flattening back the ears in a fight so as to avoid serious injury or terrorising and immobilising prey by expressing anger. It is more difficult to think of what could be, in general, the utility of the expression of fear, sadness, embarrassment, shame, hatred or envy. Should you warn your adversary? Should you announce your weakness, distress and confusion? Is it advantageous to a predator to frighten its prey? Is it useful to the prey to broadcast its imminent flight? If the struggle for survival essentially opposes individuals, should we not instead expect that selection would have favoured those best able to hide their emotions? This argument is reinforced by the thesis about human cognitive capacities put forward by those who hold the theory of Machiavellian intelligence. According to this hypothesis, the development of cognitive abilities in humans and primates in general comes from the struggle between individuals within social species. More specifically, this struggle generates the need for dissimulation, which gives rise to intelligence (Byrne & Whitten 1988).

In evolutionary biology and in sociobiology, this question has generally not been answered directly and, except in Darwin's work, the response has very often employed the concept of group selection or a related notion. The answer has not been direct because the solution to the problem of the expression of emotions is generally based on an explanation of what advantage there is in having a given emotion or behaviour rather than an explanation of the advantage gained through the fact of expressing the emotion. The explanation of the evolution of altruism using the concept of inclusive fitness is an excellent example of this. This explanation suggests how, when we take kinship relationships between organisms into account, an individual can increase its reproductive success by sacrificing itself for others, e.g., an organism that obtains a reproductive advantage by sacrificing its life to save, say, three of its children or five of its nephews. However, once we have accepted such an explanation of altruism, it is clear that there is no longer any obstacle in principle to explaining the evolution of expressive behaviour that allows an individual to obtain help from others. Explanations of the evolution of

5. Gayon (1998) has shown remarkably well that in Darwin's work the principle of natural selection must be understood as based on competition between individual organisms and that the issue of group selection is where Darwin is furthest from Wallace. See especially Chapter 2.

the expression of emotions also often use the concept of group selection.⁶ The reason for this is simple. If group selection takes precedence over selection among individual organisms, then it is no longer very difficult to see how some characteristics that are useless or even harmful to the individual, such as the expression of certain emotions, could evolve if they are advantageous to the group. For an example of such an explanation, see Wyne-Edwards (1962), and more recently Sober & Wilson (1998), for a criticism of group selection, see Williams (1966).

In this case, I think that we have to begin with the problem of the expression of emotions rather than with the question of the utility of a given emotion or even of emotions in general. In fact, as we will see clearly below, I think that *the expression of emotions* is prior to *emotions*, both logically and chronologically. In other words, I think that we should not consider what we commonly call 'the expression of emotions' as the contingent external disclosure of a prior internal state or a disposition to act, in relation to which the expression is secondary. On the contrary, *the expression of emotions* constitutes a system of communication and it is only in relation to this system of communication, I claim, that we can speak of *emotions* that we can or cannot express. Using the expression of emotions as a foundation, an explanation of the evolution of affective life can be formulated, and this explanation rests on the fact that what is in question is a process of co-ordination that is essentially and increasingly social.

3.2 The 'expression of emotion'

We can think of what is called 'the expression of emotions' in two diametrically opposed ways. The first is that of common sense, i.e., of folk psychology. From this point of view, some actions are the expression of an internal state (that we call anger or aggressiveness, for example). These actions (e.g., a change in the tone of voice, insults, narrowed eyes and changes in skin colour) should be considered warning signs preceding a form of behaviour (e.g., physical violence) or as safety valves that make it possible to avoid that behaviour. These two sub-interpretations are not incompatible and it is easy to imagine that the same actions could be both warning signs and safety valves. It should also be noted that Darwin's (1872) three explanatory principles are perfectly consistent with this description. However, in

6. They use either this concept or the notions of inclusive fitness or of kin fitness, which are, no matter what they say, related to the notion of group selection. While both inclusive and kin fitness, unlike group selection, make the individual rather than the group the one with the property of having an adaptation, all three must nevertheless be seen as means for separating the adaptive advantage from the individual. For an analysis of inclusive fitness models in terms of group selection, see Sober & Wilson (1998).

both cases the expression is seen as the very beginning of a behaviour to which it adds nothing. The unity of the whole, the link between the behaviour and the warning signs (or safety valve) is provided by the internal state that causes them both and that we call 'the emotion'. Seen in this way, the expression of the emotion is not essentially different from the behaviour that it causes. A physical attack differs from an expression of anger only in degree. They are not two different forms of behaviour but two positions on the graduated scale of the same reality. Flight is only a more extreme form of fear, which culminates in terror. From this perspective, in order to understand the existence of the expression of an emotion despite its apparent uselessness, we have only to show that the behaviour is useful in itself and its expression inevitable.⁷

The second way of conceiving of 'the expression of emotions' is quite different. It does not view expressive movements as the tiny beginnings of a given form of behaviour, but as completely different actions that have their own purposes, which can be completely different from those of the behaviour they are deemed to foretell. For example, rather than seeing aggressiveness as the commencement of a fight and fear as the intimation of flight, I think they should be seen as a system of threats and offers of submission – in other words, as promises. This is why both are something like actions, and I say 'like' because affective actions require more than the mere expression of the emotion. This suggests that for both anger and fear, a certain skill is required. The skills required are different in both cases, and they are also different from the skills that are useful for fighting, for example. The unity of the forms of behaviour that go from anger to violence is not caused by the unity of the mental state underlying them or the continuity of the feeling that accompanies them. This is only a useful illusion. There is in fact no unity, but to be effective a promise must be credible.⁸ Consequently, the spontaneous opinion

7. In fact, seen in this way, these actions limit rather than add to the behaviour in question. As safety valves, they abort the behaviour itself, and as warning signs, they undermine it because they warn the adversary or the prey. This is why Darwin's (1872) third principle applies easily in this sort of framework. The idea of direct action of the nervous system, which considers that the expression of emotion consists essentially of those movements necessarily associated with carrying out a form of behaviour, shows why movements associated with useful behaviour will be maintained despite any disadvantages they may entail.

8. What I am saying here has nothing to do with consciousness or the feeling experienced. I am not arguing for a cynical theory of anger that would say, for example, that threats are generally nothing more than hot air intended to impress the adversary and are not accompanied by the firm intention to follow through. On the contrary, I think that threats are generally sincere, in other words, I believe that anger does dispose one to later carry out violent actions. However, if it is important to take the enemy by surprise and if fraud and deception are cardinal virtues in combat, then there should be selective pressure against the expression of emotions, at least

that anger is the beginning of a form of behaviour that culminates in violence is exactly what we should expect if anger is fundamentally a threat. However, the fact that we share this belief, that for us this belief is a rule of inference and that this rule of inference is an adaptation if anger is a threat does not entail that this rule or belief is a scientifically or philosophically satisfactory description of this emotion.⁹

There is no society among wolves and lambs – the hunter simply devours its prey. The predator has no use for its prey except that which involves its prey's death. This is not true of wolves among themselves. When a conflict arises over a good of some kind, it can be to the advantage of *both adversaries* that the conflict is resolved before either suffers serious injury. The reason for this is that wolves are so similar in terms of strength and skill that there is very little probability that one could inflict serious injury on the other without being badly hurt itself. Under these conditions it is useful, as Hobbes (1651) more or less says,¹⁰ to be able to determine the issue of the conflict without undertaking the risky business of a fight. Thus, while the hunter has no particular interest in revealing its intentions before it is too late (after all, hunters who are too noisy or who talk too much come back empty handed), antagonists have every interest in informing their adversary of their intentions. Biologically, they share the same goal: that of obtaining the object of the conflict without suffering serious injury. Whatever the biological cost of an injury in relation to the advantage obtained through victory, it is not zero. Therefore, organisms that manage to secure the object of rivalry without really fighting will have an advantage. Likewise, organisms that successfully increase the probability that they will obtain the good contested without getting hurt will be more fit even if in many cases they have to back off and leave empty handed.¹¹ Conse-

if such expression is only an epiphenomenon inevitably linked to aggression. Thus the idea that the expression of an emotion, e.g., anger, is an act separate from the aggression to which it sometimes leads and not simply a preparation for combat.

9. This is an issue that is linked to the question of naturalization of mind and that merits independent investigation. The fact that an opinion or rule of inference is useful has, as is shown by the analysis of emotions, nothing to do with whether the rule or opinion is true. Attempts to naturalize the mind are mostly Darwinian or evolutionary and tend to confuse truth with advantage. As I have said elsewhere, I think that their proponents will, in consequence of this confusion, be gradually forced to abandon the notion of truth (Dumouchel 1993).

10. Given the Hobbsian adage of *homo lupi homini*, it seems legitimate to use what Hobbes says about men in order to understand wolves.

11. Only this hypothesis about the establishment of a mechanism for increasing the probability of victory but not directly related to victory can explain the evolution of organisms that, as Hobbes says (1651, p. 226), are able to distinguish between insult and damage ('injury and

quently, there will be selective pressure in favour of organisms that are disposed to establish a system of threats and offers of submission that can resolve conflicts while reducing the number of serious injuries. There will be selective pressure in favour of *the expression of certain emotions*.

The cases of anger and fear are not special, at least not if, as I suggested in (Dumouchel 1999), these terms do not repeatedly isolate stable entities but indicate salient points in a process of co-ordination. This is why Gibbard (1990) was able to develop a hypothesis fairly similar to the one I have just sketched by using anger on one hand, and guilt and shame on the other. Faced with data from the cultural analysis of emotions, Gibbard thought that his explanation was culturally relative and he therefore abandoned it. It is true that there is an obvious problem in imagining that shame and guilt play a major role in the lives of wolves. However, the problem springs essentially from the fact that we believe that the terms 'guilt' and 'shame' identify either physical states, inner feelings or even dispositions to act that have characteristic features that can be used to group them into separate categories. If, on the contrary, we take these terms to designate salient points in a process of co-ordination, it immediately becomes clear that, in some circumstances, fear, guilt and shame, as well as perhaps other emotions, share the characteristic of being offers of submission.¹² This is why the example that was just given should not be considered a hypothesis concerning the special cases of two emotions, such as anger and fear (or shame and guilt as Gibbard thought), but as an illustration of the importance of processes of co-ordination in strategic intraspecific communication. It is also an illustration of the fact, that even though emotions are relational properties, they can be established on the basis of their advantage to an individual organism alone.

damage'). While an insult causes no damage, it is a claim about the (low) probability that the one insulted will be victorious. Thus the importance of responding to insults. This shows that the establishment of such a mechanism in no way implies a reduction in the number of conflicts.

12. I am in complete agreement that we sometimes use the words fear, guilt and shame to designate things other than offers of submission. It is even fundamental for the thesis that I am defending that this be so. I think that it is because these terms designate very different things that a classification or theory of emotions cannot be built directly on them.

4. Intraspecific co-ordination and social emotions

My thesis supposes that bodily movements, which are the natural signs of certain forms of behaviour, such as fighting or fleeing in the context of predation, become the means of a process of co-ordination in strategic intraspecific communication by enabling organisms to learn about each others' behavioural options. The context of strategic interspecific communication, contrary to that of predation, exerts selective pressure in favour of successful co-ordination and therefore in favour of a certain expression of emotions. Of course, strategic intraspecific co-ordination cannot be considered as the same as the social. All forms of bisexual reproduction are examples of strategic intraspecific communication, which, for obvious reasons, plays an important role even in non-social species. However, it could be that the more numerous the occasions for strategic intraspecific communication, the more elaborate the process of co-ordination. We could even imagine a snowball effect where individuals who are best at co-ordinating their activities give birth to individuals who co-ordinate their activities even better, until success in strategic intraspecific interaction becomes the major determinant of individual fitness. Once this threshold is reached, a species becomes truly social. A species is social when the fitness of its members is determined essentially by their strategic intraspecific interactions. It is in this specific sense that emotions are social. They are both the means of success in strategic intraspecific relations and the cause of the growing importance of strategic intraspecific relations in fitness.

The example that we have just given could be misleading and be taken to imply that all of affective life is oriented toward conflict management and resolution. It should be noted that the advantage gained by co-ordination through "anger" or "fear" is not that of reducing the number of conflicts or even that of preventing more conflicts from developing into open violence than otherwise. The advantage that this form of co-ordination provides is that it sometimes enables some organisms to obtain the object of a conflict without suffering serious injury and other organisms to give up the object without being badly hurt. The fact that emotions can both lead to and be used to prevent violence is not an objection to what I am arguing. The thesis that I am proposing does not suppose that all emotions are related to conflict. It asserts that emotions exist only in the framework of strategic intraspecific co-ordination and that outside of this framework the physiological signs and behavioural dispositions generally linked with emotions are not part of affective life because in that context those events serve no purpose in terms of co-ordination.

Friendship and loyalty are examples that can help to eliminate the false impression that all emotions are related to conflict. Friendship is an emotion that has, to use De Sousa's words (1987), a singular target in the sense in which Kant

(1790) speaks of a singular judgment with respect to aesthetic judgment. In other words, friendship unites two specific individuals and in this relation each of them is irreplaceable, not necessarily in a metaphysical sense, but in the very simple sense in which the co-ordination made possible by friendship is possible only between those two people and would occur in a different manner if different people were involved. A friend is not a member of a class, unless that class is a singleton, and is not replaceable by anyone who fulfils an equivalent description. This means that friendship enables co-ordination between people rather than between specific actions. Friendship establishes between people dispositions that are extremely useful for dealing with the unexpected. It ensures unconditional reciprocal support in a wide range of situations. It is an alliance that promotes the pursuit of common goals and complicity that facilitates the success of activities requiring the simultaneous implementation of different skills. It is one of the means by which we determine our preferences. To begin with, we like what our friends like and we allow their choices to guide our own. Tell me who is your friend and I will tell you who you are, says the proverb. Our successes and our failures are not independent of our friendships. Friendship is part of a process of co-ordination that is indispensable to our lives.

5. Emotions, co-ordination and agency

Emotions are not the only possible form of intraspecific co-ordination. Some social species, such as ants, bees and termites, use completely different mechanisms of co-ordination. What seems to be characteristic of affective life as a process of co-ordination is that it constitutes the very agents that it co-ordinates. The co-ordination produced by affective life does not consist in adjusting independent actions to each other but in linking independent agents. Yet, affective life is the means by which agents become independent. Emotions as a process of co-ordination are the means of reciprocal influence among people and this influence enables individuals to acquire some autonomy. In other words, emotions are the tie by which we separate ourselves from each other. There is no independent subject or autonomous individual before the affective process that connects us. Emotions constitute us as subjects. They are the process that makes the relation to the other prior to autonomy. That, at least, is the thesis I now want to defend. The idea that the process of co-ordination pre-exists the agents that it co-ordinates is not as paradoxical as it seems at first sight. In its most abstract form, it means that the continuum is prior to the discrete.

I said above that the expression of emotions is a process that enables co-ordination among organisms because it is how they reciprocally inform each other of

their behavioural preferences, e.g., that they prefer fleeing to fighting, fighting to mating or rejection to agreement. This formulation lends itself to confusion because it implies that there are well-defined preferences and set forms of behaviour of which organisms inform each other. Affective life must instead be seen as the process by which these preferences are established and these forms of behaviour determined. In this way, it is a mechanism that both co-ordinates and constitutes agents. While fear, anger and love can sometimes be considered points in affective life in which orders of preferences already exist and forms of behaviour are already determined, affective life itself is the process by which these preferences are established. Melancholy, happiness and anxiety are generally better seen as points at which preferences are not set and behaviour is undetermined. Being in love, at least in a certain sense, is a classical example of a relation in which the preferences of the lover vary in accordance with those of the beloved. In other words, it is a relationship in which at least one of the parties does not arrive with a set list of preferences or forms of behaviour, but leaves it up to the other party to establish that list. The emotion and proclamations of love can therefore be seen as the commitment and promise to abandon to the other the responsibility for setting our preferences. The beloved's indifference and the lover's lack of independence form major themes in literature and popular psychology. Specialists in interpersonal relations either take the lover's side and chastise the beloved because the lover is the only one in love, or they reproach the lover for a lack of independence, indecisiveness or 'weakness'. In the former case the beloved is criticized for a lack of reciprocity, in other words, for an inability to allow his or her preferences to be determined by those of the other. In the latter case, the lover is criticized for a lack of initiative, in other words for an inability to determine the preferences of the other. However, in both cases, what is lacking is the reciprocal determination of preferences and behaviour. In contrast, in a less distorted relationship of love, the emotions of each party are not independent of a disposition to set one's preferences in accordance with the preferences of the other. This reciprocal disposition shows clearly that the preferences that form the basis for our actions are not the origin, but the result of the process of co-ordination.

It should be noted that the co-ordination involved in affective life consists more in being on the same wavelength than in adjusting one's actions to those of another, for example, in order to row together. Indeed, most actions in everyday life require some degree of co-operation from the other in order to be successful. This is because most of the things we do in our lives as workers, lovers, members of a family, friends, consumers and providers of services involve others. It is rare that our decision to go to the movie theatre, park our car in an underground parking lot, go home right away, try on yet another piece of clothing before making a choice, put an important task off until tomorrow, cross on a red light or stop at the

corner store before going home is independent of the influence of others. This is simply because our actions require something from the other in order to be successful. They require that the other perform an action in return, or at least show a degree of indifference that allows us to perform the action. That this is so simply means that we are social beings who constantly interact with others and who almost always act in their presence. Emotions are related to the skills required to carry out these many actions and interactions. They are salient points in the open process of co-ordination that enables us to interact reciprocally.

Moreover, the process of interaction is what makes it possible for us to act. Allow me to explain. In a context such as that of a family, the influence on others that is exerted by an individual's decision is often very visible and immediate. The links of interdependency between family members are close enough for any individual action to potentially have major repercussions on the lives of the other family members. Routine is the most common solution to this problem. It consists in each member repeatedly performing certain actions at set times and in this being known so that these fixed points of action enable the other family members to co-ordinate their activities. Traditionally, this solution was reinforced by another answer to the problem, which was to assign to a single person, the father of the family, the power to make decisions that could affect the lives of the others. Neither the mother nor the children were supposed to make decisions with respect to actions that would take them away from the family routine. Indeed, except in extraordinary circumstances, they were not in situations in which they had the chance to do so. In such conditions of extremely tight inter-relations, outside of what is determined by routine or authority, my initiative to perform an action, even if it concerns only myself, such as drink a glass of water, is always a potential object of agreement or disagreement. "Don't drink so much", they may say, "it will make you sick" or "it ruins your appetite", or else, "drink, drink, it's good for your health". My initiatives, incoherent seeds of imprecise actions, become decisions in relation to others. It is in relation to agreement or disagreement with others that I progressively acquire the ability to decide, partly because I come to see that the distribution of agreement and disagreement is not purely random.

One might ask how this is a process of co-ordination. Agreement or disagreement, or a certain indifferent presence, are like promises in that they authorize expectations. More specifically, the expression of agreement or disagreement reveals a disposition to various forms of behaviour. It is not the indication of a specific form of behaviour but rather of something like what Hayek calls 'negative rules'. Disagreement does not signify any particular action; it does not mean attack, withdrawal, rejection, anger or flight. However, it excludes, or at least makes improbable, a certain set of forms of behaviour. This disposition can nevertheless immediately be seen as the manifestation of an order of preferences, in so far as

agreement or indifference are options for an agent moved to disagreement. However, and this is important, the agent's 'expectation' upon perception of disagreement remains so undetermined that he or she is unable to anticipate clearly the future behaviour of the person with whom he or she is interacting. Saying that expressing agreement or disagreement is like making a promise able to produce an expectation, does not suppose anything more from the person exposed to the expressed emotion than his or her own emotional reaction. That person's resistance, questioning, dissent, anger, fear or conciliation is his or her expectation and promise. Not taking action, helping, or opposing, form the content of that 'expectation'. Generally, because our actions so often take place in an environment in which others play fundamental roles, no one can carry out an action successfully independently of the affective reactions of others that fulfil or contradict his or her original expectations. The responses of the other determine my strategies and, in return, my expressions of emotion allow the other to identify the framework in which these strategies can be evaluated. Co-ordination results from a relation between individuals that has precedence over their relation to things. Agents determine their behaviour in function of each other and determine their relationships to things in function of their reciprocal relationships. Their co-ordination is not the adjustment of two actions but the establishment of a framework relationship defined, for example, by a disposition to conflict, co-operate or be indifferent and, as I tried to show in (Dumouchel 1999), it is only within such a framework relationship that it makes sense to choose between strategies. The choice between different strategies, in other words, between specific actions leading to results that can be evaluated, is only possible within an already predetermined framework that specifies the game or games we propose to play. Our affective life is made of the ups and downs of the process by which we co-ordinate ourselves with each other, and we call the salient points of this process emotions.

The fact that faced with your disagreement, I adopt a new approach in order to achieve my goal is often called determination, a term which is not, in this circumstance, without affective value. We also very often look to emotions to explain the fact that, when you disagree, I no longer want what brought us into conflict. As Hobbes noted, depending on a person's relationship with someone who has changed his or her mind, that person will call the change in preferences 'fear', 'friendship' and 'loyalty' or 'fawning'. No matter what it is called, the affective relation indicates a means by which the actors co-ordinate their actions or, more precisely, co-ordinate with each other.

Now, it may be asked, what can be said in the context of an emotion such as mourning, the very special sadness that we feel when a loved one dies or is far away? Here the question is not with whom (the soul of the departed?) mourning allows me to co-ordinate but whether the disappearance of a person can consti-

tute a salient point in a process of co-ordination. This seems to pose no particular difficulty. If the essential purpose of affective life is to enable co-ordination between agents, i.e., co-ordination in which the relationship between the agents has precedence over the agents' relationships with things and over the agents' specific actions, then the disappearance of someone dear certainly can be a salient point in the relationship of co-ordination.

6. Conclusion

The social emotions thesis does not claim that society is based on emotions, love or sympathy, in opposition to reason, for example. It is not a variation on the theories of moral sentiments. I do not claim that the social order depends on the presence of certain feelings, nationalism, compassion or the spirit of competition. In this chapter, I have argued:

1. First, that strategic intraspecific co-ordination is the area *par excellence* (although not the only, as e.g. domestication of animals also establishes a fairly similar, although less complicated process) where selective pressure can be exerted in favour of a process of co-ordination among agents.
2. Second, that the expression of emotions can be understood in the framework of the evolution of such a process. One of this thesis's major corollaries, which has already been mentioned and which I analyse in detail Chapter 4 of (Dumouchel 1999), is that the *emotion* is secondary to *the expression of the emotion*.

One might ask why this phenomenon should be called 'social' rather than 'inter-individual' or 'communicational'. The answer to this is that, on one hand, it is at the foundation of what we more commonly designate by the term 'social' and, on the other hand, it is a very special form of the inter-individual or communicational. Let us begin with the latter point. As I discuss in detail in Chapter 6 of (Dumouchel 1999) and again in (Dumouchel 2007) affective co-ordination is based on mechanisms of a special kind, which are quite different from those we assume are at work in most models of inter-individual interactions. If these very frequently postulated mechanisms exist, then affective co-ordination is a special phenomenon that deserves its own name. The word 'social' seems to be just right because it suggests that the mechanism of intraspecific co-ordination is the source of the set of phenomena that we call 'social'.

Emotions are social first of all because they constitute us as social beings, and this they do in two senses:

1. They make us sensitive to each other; i.e., they bring us into relation with each other and allow us to co-ordinate with each other.
2. In and through these relations, they make us able to choose. This is because, without co-ordination with others, situations remain too undetermined for choice to be possible and because every affective choice includes the other.

We are agents whose individual ability to choose is social because it depends on emotion in two ways. Emotions are also social in that they provide the matter on which social rules and organizations are erected. These rules and organizations are meaningful only through the role that strategic intraspecific co-ordination plays among us. Affective life produces the interest that we have in each other. It is made up of our decisions to dominate, flee, submit, unite with the other, make reciprocal commitments, understand each other, fight, separate, be indifferent, ally ourselves, watch, fear, challenge and comply. Social organizations offer codes and rules concerning all these decisions. They tell us who to respect, flee or pursue, when to obey and when to be tender. These codes influence the tone of our affective life, but they are never identical with it. They mould our emotions but they do not construct them any more than they can be deduced from them. Our emotions are the matter that these codes regulate.

Finally, emotions are social because affective actions exist in a very special way that requires a relationship with the other. Affective actions are acts of co-ordination and, consequently, they are such that they cannot be carried out alone by the person who initiates them. Emotions are social because they always supervene over more than a single agent.

Acknowledgements

I wish to thank Lola Cañamero for her help and edition of this chapter.

References

- Byrne, R. & Whiten, A. (Eds.) (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford: Clarendon Press.
- Darwin, C. R. (1872/1965). *The expression of emotions in man and animals*. Chicago: The University of Chicago Press, 1965.
- de Sousa, R. (1987). *The rationality of emotions*. Cambridge, MA: The MIT Press.
- Dumouchel, P. (2007). Biological Modules and Emotions. *Canadian Journal of Philosophy Supplementary Volume* 32, 115–134.

- Dumouchel, P. (1999). *Emotions: essai sur le corps et le social*. Le Plessis-Robinson: Institut Synthélabo pour le progrès de la connaissance / PUF.
- Dumouchel, P. (1993). Ce que l'on peut apprendre au sujet des chauves-souris à l'aide d'une télécouleur. *Dialogue*, 32, 493–505.
- Gauchet, M. (1992). *L'inconscient cérébral*. Paris: Seuil.
- Gayon, J. (1998). *Darwinism's struggle for survival: Heredity and the hypothesis of natural selection*. Cambridge: Cambridge University Press.
- Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgements*. Oxford: Clarendon Press.
- Hayek, F. (1988). *The fatal conceit: The errors of socialism*. Chicago: The University of Chicago Press.
- Hayek, F. (1973). *Law, legislation and liberty*. [3 Vols.]. Chicago: The University of Chicago Press.
- Hayek, F. (1967). *Essays in philosophy, politics and economics*. Chicago: The University of Chicago Press.
- Hobbes, T. (1651/1968). *Leviathan*. (C. B. Macpherson, Ed.). Harmondsworth: Penguin Books, 1968.
- Kant, I. (1790/2000). *Critique of the power of judgement*. (P. Guyer & E. Matthew, Trs.). Cambridge: Cambridge University Press, 2000.
- Kant, I. (1784/1993). The idea of a universal history in a cosmopolitical view. In *Essays and treatises* (pp. 409–432). Bristol: Thoemes Press, 1993.
- Mead, G. H. (1934/1974). *Mind, self and society from the standpoint of a social behaviorist*. (C. W. Morris, Ed.). Chicago: The University of Chicago Press, 1974.
- Ross, D. & Dumouchel, P. (2004). Emotions as Strategic Signals. *Rationality and Society* 16(3), 251–286.
- Sober, E. & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton: Princeton University Press.
- Wynne-Edwards, V. C. (1962). *Animal dispersion in relation to social behaviour*. London: Oliver and Boyd.

CHAPTER 2

Fabricating fictions using social role

Lynne Hall and Simon Oram

1. Introduction

Existing within any human culture demands consideration of human interactivity, mutual engagement and reflections on our collective representations. As humans we are not hermetically sealed individuals, we learn about our world and how to operate within it from others in a myriad of ways. Embodied agents operate within culture; they “work” in a particular situation carrying out a particular role. It is the aim of our research to construct an agent that is “aware” of its social situation, that is, the agent has the ability to emit signs that contextualise it within a real world setting.

Culture, unlike a written narrative, is transitory: it may be the case that once the present has receded into the past there is a level of narrative fixity (Event A happened which lead to event B), but the present is still an open ended play of improvisation (Layton 1997). This is not to say that individuals have an unbridled ability to think in an infinite number of different ways, from an infinite number of different positions. Human action, ways of communicating and thinking are limited by particular historical and spatial constraints, that of living in a particular culture at a given moment in time. Given these conditions, how we articulate this in an agent’s behaviour is, to say the very least, problematic.

Current technology doesn’t have the ability to produce cultural complexity in any convincing way; passing the Turing test looks further away from being realised now than it did when it was first proposed. Firstly, quite simply, present programming paradigms cannot produce simulations of cultural behaviour in any believable way (Dreyfus 1992). Secondly, to many observers, the humanities have reached a crisis of representation, that is a “science” or meta-theory of the social is not achievable (Spiro 1996).

On one hand there a need to create socially able individuals, because they are more believable and natural human/computer interfaces. But on the other there are criticisms as to whether this is achievable. Given this, what hope is there for social-ability?

Being able to design an agent that can fully negotiate its social surroundings (however desirable) is not what we aim to achieve. Rather we want the agent to have access to a narrative of its own cultural context, in our particular case this is the organization of meetings within an office environment. The agent will be able to evaluate its environment in a rudimentary way and through this process achieve a level of believability. The agent's social role in relation to the meeting participants will provide the driving force for the interaction. This social relationship is based on the application of Foucauldian ideas where the key contextual variables are determined through power relationships realised within the organisational hierarchy.

Section 2 considers issues emerging from recent work on social ability and animated agents that have had an impact on our approach. Section 3 discusses a Foucauldian approach to fabricating fictions through the creation of a fictive personality as an embodied agent. Section 4 describes PACEO, an agent that appears to have an awareness of social context through an understanding of power relationships. Section 5 presents a brief evaluation of PACEO. Section 6 discusses findings and indicates directions for future work.

2. Agents and social awareness

The aim of our work is to create an agent that enables us to explore social interaction between agent and user. This social interaction will be constructed through the agent's awareness of its own and other users' social role. Awareness of social role will mean that an agent will be able to adapt its behaviour depending on whom it is communicating with.

There is growing interest in the use of social role (Prada & Paiva 2005; Rist & Schmitt 2002, and this volume) as a way to support the development of appropriate interaction patterns (André 1999). Hayes-Roth (1998) studied role-specific interaction from the position of a master-servant relationship, with agent positions that were essentially fixed. Believability was achieved by drawing from drama and literature, with the agent having a specific, narrow, defined role.

The agents of the Oz project (Bates 1994; Mateas 1997) also exist within a fictive environment, appearing to have social awareness and attempting to create an illusion of life. The agents have social abilities that evolve through interactions; however, these abilities are defined by the characters' role within the plot. The agents exist within fictional, dramatic environments separated from the context within which they are used. The agent that we are developing will have a specific, narrow, defined role, however: office life rather than drama or fiction provides the context.

Prendergast and Ishizuka (2001) use social role to enhance the believability of an agent, basing the adjustment of behaviour on the applicable role in a socio-organisational structure. Users interact as conversational partners within a role-playing language-learning environment. We also intend to use a socio-organisational structure to determine action and behaviour; however, this will be based on notions from Foucauldian power structures and will be focused at a real world rather than a role-playing environment.

Similar to Rist (Rist & Schmitt 2002, and also in this volume) our intention is to consider the social relationship of agents and users within the problem domain of meeting organization. Rist's agents negotiate in situations where the outcome of that negotiation is not explained only through rational argumentation but also in light of the social context and the personalities of the negotiating partners. Users are represented by their agents, who inherit their users' social and contextual relationships. These agents follow a negotiation process driven by their personalities, moods and emotions and the social relationships amongst them as indicated by their owners. In our work, we have only a single agent who interacts with many users, thus our focus is on interaction with users rather than with other agents.

The approach we have taken is similar in some aspects to the Oz project (Bates 1994; Mateas 1997) where, rather than attempting to construct an "intelligent" piece of software, a more pragmatic view is taken. What is most important for Oz is believability, it is not concerned with the internal workings of the agent, but rather how the agent engages its audience. How an audience comes to believe in the agent is central to the Oz philosophy and essentially central to our own.

3. Constructing the agent's culture

Where Oz stays with fictive dramas, we want the agent to be immersed within the narratives of everyday life (Polaine 2005). We will approach this by drawing from ethnography and social theory to build a narrative of the social space that the agent will function in.

In a way we can see the agent as a kind of ethnography. By ethnography we mean a textual rendering of an area of culture/society. This is usually achieved by a protracted time "in the field" with the "natives" documenting their lives. The ethnographic raw material is then delineated into some form of theoretical/structured framework that becomes the ethnography. Ethnography collects specific information about specific situations or events and as so could be very useful for creating situated agent narratives within real world settings.

But it is not simply the case that once transported into the "field" the ethnographer can collect data in some form of value-free way. Cultural baggage and a

methodological agenda create a rendering of the social that says nearly as much about the ethnographer as it does about those under study (Clifford & Marcus 1986). To an extent ethnographies of Western institutions can circumvent some of these situations, especially when they are turned back on the institution that the ethnographer is situated within.

What is different in this project is that the “field” is the workplace within which the ethnographer operates. The notion of ethnographer as subaltern solves some problems of representation, such as the having an “authentic voice”, as it is essentially a work of autobiography. But it still fails to sort out the issues of methodological *a priori*.

There is a need then to be explicit about the methodological tools that will be used to guide the ethnographic endeavour. In our case, the methodological basis is the Foucauldian method. It is used as the roadmap for writing the ethnography, which will in turn be used to inform the design of the agent.

3.1 A Foucauldian approach

Although Foucault is primarily a historian of systems of thought, his main aim is to uncover the conditions that give rise to the socio-cultural peculiarities of contemporary Western institutions. Foucault achieves this by critically tracing through his method the historical emergence of discourses on sexuality, medicine, race, class and how these discourses interplay with concrete institutions such as the workplace.

3.2 Foucault’s notion of discourse

Foucault’s notion of discourse is not one of discovering general underlying rules of language that lie “behind” texts. Neither is it an attempt to build formal linguistic systems, their dynamics or uncover what the mechanics of argument are. Rather, Foucault is interested in the cultural conditions that allow statements to be considered true or false (McHoul & Grace 1997). The statement for Foucault is the unit of discourse just as the sentence is the unit of language.

Foucault’s theory of discourse asks – what are the set of rules that allow us to say or think something at a particular historical or cultural point? For Foucault, discourse is a technical term that denotes relatively well bounded areas of social knowledge, for example criminology and sexuality (McHoul & Grace 1997).

Foucault’s fundamental premise is that an individuals’ self and sense of the world is constituted through discourse. His theories move away from the structuralist approach which attempts to discover the universal and objective constituents

of human thought patterns, and the hermeneutic or phenomenological approach that places the individual as the primary arbiter of how they interpret what they see. Foucault pluralizes the structuralist position and removes the primacy of the individual in the interpretative task by de-centering the self. On a simple level Foucault's view of the self can only be the sum of ideas about the self that an individual has access to; these ideas have pre-figured the individual (Gutting 1994).

Discourses are constituted by the dynamic interplay of power and knowledge. This allows a more subtle and dynamic theory of power than previous attempts such as Marxism in which one group has an intrinsic hold on power over another group (Gutting 1994). It is knowledge that gives power its validity, but also power allows knowledge to perpetuate, by giving it hegemony.

3.3 Bio-power

The discourses that characterise western/contemporary culture and allow it to function are for Foucault bio-political. Bio-power constitutes a set of techniques that render an individual docile and productive. Bio-power effects operate at the level of the population but more importantly for us they operate on the micro-physics of concrete everyday social interactions (Rabinow 1984). Such bio-political discourses can be uncovered within the workplace (Mitchell 1999).

The general mechanism of bio-power for Foucault is one of discipline (Foucault 1977). This is realised as both the disciplinary surveillance of an individual, by those "in power", or more subtly as the construction of subjectivity through an individual's self-activity; the constant surveying of the self so as to effect a change in behaviour in response to disciplinary regimes at work within institutions.

Discipline works in four major ways. Firstly it employs dividing practices, that is, the social and sometimes spatial separation of individuals within a population. Secondly, control of activities is brought into place. Its focus is the extraction of "time" from bodies: the daily timetable, clocking in and out of work; by adjusting movements, such as marching, and good handwriting; and by articulating the movements of the body with an object such as a computer. Thirdly, discipline is concerned with stages of training, the development of an individual's ability, by way of pedagogical practice, in a process of movement from student to master. It codifies these stages within terms of hierarchy, increasing the level of difficulty of each stage. This enables the ability to monitor skills while also providing a way of separating and individualising. Finally, discipline also brings into effect the coordination of these parts to place the individual into the general machinery of the institution they are situated within.

3.4 Foucault and the workplace

The workplace is an institution where these disciplinary activities are carried out on a daily basis. It is practices like these that are realised within discourse at the level of statement. Our ethnography presently focuses on one general area of discipline: that of dividing practices. Dividing practices work by the categorisation of individuals at binary opposites, for example the sane/insane, straight/gay, employed/unemployed (Foucault 1977).

The workplace itself is a good example of a dividing practice. As an institution it physically brackets the worker from other parts of life. There are also divisions through the institution with discourse generating the very fabric of power relations that demarcate manager from worker, educated from manual worker, good employee from poor employee, trainee from the rest of the workforce (Foucault 1977).

While discourse constitutes possible discursivities, where there is power there is also resistance and counter discourses. Counter discourses do not just operate in grand abstractions, they operate on the micro level of everyday life, such as justifications for better pay. Discourse defines the objects of analysis but also defines the field in which these objects are talked about, that is, the rules of engagement for both the statement and the counterstatement.

3.5 Developing the ethnography using Foucault's method

Dividing practices thus define the demarcation of individuals within the workplace and they are realised as concrete relations and form a quite simple taxonomy. Within these dividing practices lie many salient narratives of the workplace; it is these narratives that we want to document in our ethnography. This gives narratives that are local and specific, and ones that can be articulated within a general taxonomical framework of power relations.

An ethnographic analysis has been carried out on a small research team, within which both authors are situated. The ethnographic raw material has been gathered over a period of ten months, through participant observation (by the very fact of working in the team), and by the collection of narratives from individuals about their daily experiences: these may have occurred at the level of a question posed by the ethnographer or as part of a general conversation within the office. The ethnographer has focused on how these narratives are constituted within the intrinsic power relations within the organisation.

Foucault's framework then gives us a basic theoretical yardstick to both delineate the social it into ethnographic description as well as give us a rubric for

social variables from which to hang the agent's utterances. It is hoped that these utterances because of both their cultural specificity and dynamic nature are believable to the user.

4. Fabricating fictions (creating the illusion of social ability)

The agent's view and the user's view of the agent are constructed by the discourse within which they co-exist. Agents may have access to a very limited sum of the self, however, due to their existence as fictive personalities they may appear to be considerably more rounded than is actually the case. Incorporating agents into a workplace culture may result in the agent being ascribed attributes that it does not possess but that would be in line with the complex view of self that users have.

4.1 Situating the agent

The socio-organisational structures typically considered, like that of Prenderinger, and Ishizuka (2001) and Paiva and Prada (2005) are within recreational environments, where there are no explicit power hierarchies such as that seen within most working environments. Here, the intention is to place the agent at the mid-point of an organizational hierarchy, see Figure 1, with the specific task of organizing meetings. The agent is entitled PACEO (Personal Assistant to the Chief Executive Officer) and is under development to organise meetings within a workplace context.

The target user population is the hi-tech, white-collar sector, who has the computer continuously available. The aim was to create an agent that convinced users that they were interacting with an agent with social awareness rather than actually creating an agent that had social awareness. It is not the 'personality' that is to be developed, but rather the portrayal of a fictive personality.

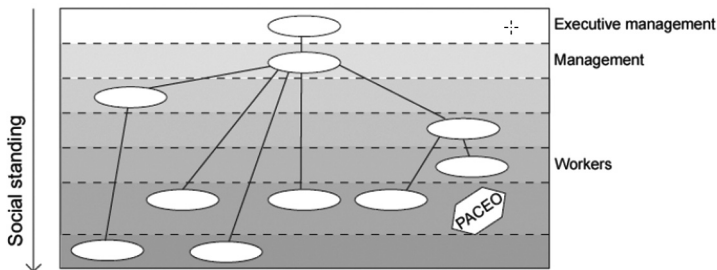


Figure 1. PACEO's organigram

4.2 Developing the agent

PACEO was developed with Microsoft Agent, using one of the standard, freely available animations. It interacts with the users within the context of a diary application similar to other such time-management systems, i.e. Outlook. PACEO 'inhabits' a small LAN and roamed its domain searching for agreement with the various meeting participants.

4.3 Fabricating the agent

The ethnographic/social analysis of the workplace gives us an insight into the cultural working of the office environment. It is from this analysis that the agent was constructed. The agent is not based on any one individual within the organisation, but it is given its own unique position and role, tied to its function within the office that of organising meetings. The character of PACEO is purely fictional. We wanted an agent that had a stereotypical character, one of a white, male, thirty-something.

For example PACEO is rather 'old fashioned' in his views of women (Coates 1994), and finds interaction with them problematic. Because of this PACEO resorts to using language that people may find slightly too intimate within what can be considered "correct" behaviour within the office. He is also rather conservative, which realises itself in the office as siding with those in power, seeing the discourse on hierarchy in the organisation and its position within it as the main guiding force for social interaction. So although PACEO is fictitious, he is fiction based on the effects of disciplinary practices that we have delineated within our social research. From this context, language is drawn from which the agent uses in its interactions with users, including passive acceptance of role, sycophancy towards those in power, dismissal of those lower in the hierarchy, etc.

PACEO is portrayed as an individual, autonomous entity to the users. It does not appear to reside on all the user's machine, rather it imitates the behaviour of a mobile assistant, appearing only when requested or when sent to organise a meeting by another user. Thus, when the agent needs to interact with the user, it graphically appears on their screen and then 'leaves.' The activity and dormancy of the agents was controlled centrally, with user events (such as request for meeting) affecting agent status. This approach was taken with the intention that users should feel that the agent was always somewhere on the system either carrying out its roles or waiting for its next orders.

4.4 Fabricating the users

For PACEO, each user “appears” as a set of variables. These variables are used as a basis for differentiation between users, allowing PACEO to adapt its utterances and behaviours depending upon whom it is interacting with. The interaction with the agent is determined from the social standing of the user, the agent’s current mood (which is dependant on recent interaction history), user manipulation of the interaction context (selecting an attitude towards the agent, which is somewhat crudely provided as love / neutral / hate) and social relationship (derived from the previous relationship that the agent has had with the user).

4.5 Social standing and social distance

Social standing is determined from considering the individual’s position within the organisation. This is drawn from the earlier ethnographic work. On one level this was a simple case of creating a hierarchy based on job title of all the individuals in the team. However job title was not the only determinant factor. Social distance also played a quite fundamental role in how one was positioned within the discourse of the office.

This could be seen quite clearly in the flow of conversation: who chatted to whom, how language (especially the use of academic and scientific discourse) clearly had the effect of demarcating academic from non academic employees. It could also be seen in who went to lunch together and who socialised outside the office. This is not hierarchical in the sense of role and position, rather this seems to be operating as a set of in groups and out groups that had a level of fixity over time.

Initially, when the user requests that PACEO should help them to organise a meeting, PACEO will determine the social standing and social distance of the user in relation to itself and select the appropriate behaviour. In the actions used to inform participants of the meeting that there is a call for the meeting, the utterances selected are based on the position of all parties involved in the meeting. This includes the sender, the recipient and the agent itself.

PACEO has a map of the organisational social standing representing its view of the organization, see Figure 1. This effectively provides it with a level of respect for each social interaction. PACEO’s view of the organisation will change over time as it develops relationships with the various members of staff. The social standing of staff will also change within permitted boundaries, for example PACEO may have a poor relationship with a manager (high social distance), however, due to the manager’s position within the power structure (above that of PACEO), the utterances selected may be cold and formal, but they will also show the subservient position of PACEO.

4.6 PACEO's mood: Relating to users

The variables relationship to user and last time seen, express the social interaction between user and agent. PACEO has a 'memory' of the users (Prendinger & Ishizuka 2002; Ho & Watson 2006), remembering prior interactions with them in terms of when the interaction occurred and the user's treatment of the agent. This historical relationship is critical to PACEO, as its social interaction with users is based on social variables, previous user relations and the agent's current 'mood.' Any user interaction could potentially alter its mood and subsequent interactions.

The user could interact with the agent neutrally or could adopt one of two moods in conversing with the agent: love and hate. Selection of either of these (presented as buttons in the response dialog box) would influence the general mood of the agent as well as increasing or decreasing its relationship to the user. Interacting with the agent consists of five movements: appearance, introduction, work, farewell and disappearance. Each one of these areas consists of a variety of utterances depending on the mood, the agent's memory of the user and the impact of the social variables.

The social variables that represent the user are filtered through the agent's mood. The mood manipulates the social relationship and social standing variables by either increasing or decreasing them. This has the effect of creating a more agreeable / disagreeable agent both for the subsequent user and for the current user's subsequent use of the agent. Being unpleasant to the agent (selecting the hate mood) or using a range of unpleasant utterances tends to result in the agent being less pleasant with subsequent users and in any further interactions for the offending user.

Thus we see that earlier interactions, both with the current user and with previous participants affects the interaction approach that PACEO will take. In this sense the agent will be unpredictable, with some users (particularly those who are in the same or lower organisational position than PACEO) getting different treatment. However, consistency in terms of modes of interaction and task structures will be maintained, thus although the user will not know the mood of PACEO prior to its arrival, they will be able to predict the outcomes of the interactions, such as booking a meeting, agreeing to participate and updating the diary. This inconsistency in terms of mood is seen within the utterances used and early testing with PACEO suggests that part of the desire to interact with PACEO is based on wondering what it will say.

5. Evaluating PACEO

The research question we were trying to evaluate was whether with PACEO we had created a simple portrayal of a believable fictive personality, based on Foucauldian notions. Evaluating this question was achieved through a qualitative evaluation approach where the user interacts with the system with a domain expert (researcher) constantly available and attempts to perform guided or pre-scripted tasks. The user is encouraged to make comments throughout the session and a debriefing is used to summarise activity. The evaluation was conducted with 8 users, who had similar characteristics to the intended target user population (e.g. highly computer literate at the end user level, with specialist abilities in specific products).

5.1 Procedure

Each of the sessions involved a single user and researcher with users interacting with PACEO for approximately two hours. At the beginning of the session an overview of the purpose of the evaluation was given, identifying that each of the users had a role (equivalent to organisational position) to play in a fictitious organisation, where meetings were organised by PACEO, an animated agent.

The session began with an overview of PACEO and the diary system, involving approximately 10 minutes of task-based use of PACEO. During the session, the users spent their time on personal computer-based work tasks (i.e. writing reports, preparing a lecture, etc.) simulating a work environment. Quality scripts provided the details relating to meeting organization involving the user in each of PACEO's main roles, with the user exploring meeting initiation, rejection and acceptance, see Figure 2.

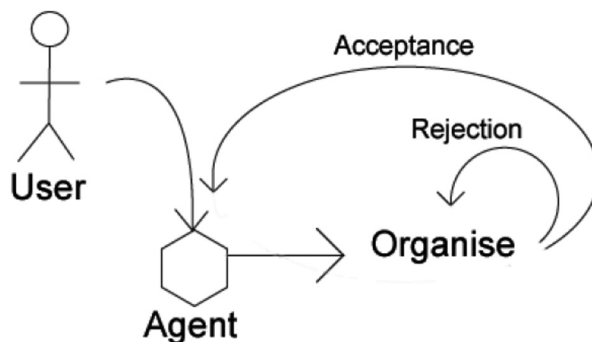


Figure 2. PACEO's roles and interactions with users

After the session users were debriefed, being asked for their opinions of PACEO. The main focus of these questions was on their perceptions of PACEO's believability and its ability to engage with them. Users were also asked about their satisfaction and enjoyment in using PACEO and interest in using it again.

5.2 PACEO's usability

None of the users had significant usability problems and were able to use PACEO to accept / initiate / reject meetings. The diary system upon which PACEO operates was found to be acceptable and overall the system was judged to be competent and as participant F noted: *"fun and a lot more interesting than Outlook."*

Users who found PACEO amusing were more likely to be positive about it. For example Participant F, to whom PACEO was fairly chatty said: *"Yep, it's easy to use and I like the interactivity, It may even liven up humdrum office chores, as long as it didn't get in the way of work in the office, er, which I don't think it will."* However, Participant F had found the agent highly engaging and amusing: *"I think the humour lies in the fact that computers shouldn't be like this, have a personality of their own."* Meanwhile, Participant B who was uncertain of whether his requests would be carried out was understandably less enthusiastic about continued use of PACEO, although he did concede that he would use it again.

PACEO was exceptionally rude to Participant H, as the following statement identified: *"My only concern is that he may rub people up the wrong way, may say something that isn't acceptable [as he did to Participant H], or for instance if you have had a bad day they last thing you want to hear is your bloody computer telling you off and being nasty."*

Participant H had the most negative experience of all of the participants and she had little desire to continue to use PACEO, criticising it for limited utility rather than agent aspects: *"I don't know, I feel that this program may get in the way, what is wrong with organising meetings by email, telephone face to face, rather than needing a computer program to organise things for you, all I'm trying to say is what utility does it offer?"* Her preference for dealing with people rather than software is apparent in her comments, however, it could be suggested that her preference is informed through a negative experience with PACEO and her stated dislike of it.

Where PACEO exhibited relatively neutral behaviour, users found it less engaging and interesting to interact with, typically finding it *"dull."*

Table 1. User social variables and relationship

User	Age	Org. Status	Gender	Relationship
A	50	High	Female	Friendly
B	20	Low	Female	Unfriendly
C	35	High	Male	Friendly
D	30	Medium	Male	Friendly
E	35	High	Male	Friendly
F	25	Low	Female	Unfriendly
G	58	Medium	Male	Ambivalent
H	30	Medium	Female	Unfriendly

5.3 Social role: Does PACEO (appear to) understand power?

Table 1 provides the social variables of each of the participants and PACEO's most typical social relationship with each user.

In the case of Participants A and E, the agent was extremely subservient, responding using utterances that highlight the hierarchical difference in positions between itself and the user. Participant A commented "*He's very crawly isn't he! ... [but] ... quite amusing.*" Participant E, the CEO of the company and thus at the top of PACEO's horizons, stated: "*... I like the way he is on a certain level aware of you, it's quite amusing that he is sucks up to me in my role as a CEO*" This participant had an enjoyable interaction throughout, with PACEO at his most sycophantic. Participant C also dealt with the sycophantic agent, however, his response was "*I don't really like this character, he's a bit slimy isn't he. Even seems a little sleazy...*" Participant D had a similar social body to PACEO, but the participant, who had a friendly relationship with PACEO found the agent boring "*Well, he's bit dull towards his friends isn't he? Would be good if he knew office gossip...*"

5.4 Social engagement with PACEO

As the user only interacted with PACEO for a limited session, the interaction history was initially pre-set and then modified during the interactions with the user. The user could manipulate the love / hate buttons to give the agent greater awareness of the user's perceptions of it. However, even in a limited time period, the evolution of a social relationship could be viewed. For example, the evolution of the interaction between Participant B and the agent showed a steady deterioration of the relationship, sufficiently so as to affect Participant B's trust in PACEO.

Many of the users talked about the humorous aspects of PACEO. However, Participant B took a 'hate' approach to PACEO, finding the increasing rudeness

of the agent highly entertaining: “*What a cheek... [laughter] ... Never been talked to like this by a computer before!*” Clearly, however, the agent’s negative approach had an impact on the user who also stated “*The way it was talking to me, I wonder if it will actually book this meeting...*” This is a disturbing result, for whilst amusing as a one-off experience within what is effectively a gaming situation, this may be very stressful within a work environment. It could also have a negative impact on the task, with the user feeling the need to confirm the meeting by some other means (face-to-face, email, etc.).

All of the users anthropomorphised the agent, referring to it as he and assigning characteristics and attributes to the agent, that it didn’t possess. Participant D received a visit from PACEO after it had had a negative interaction (with Participant B) and stated, “*To be honest, I don’t think he likes me, mind you, he probably doesn’t like anyone.*” Participant A gave her view of PACEO: “*You know he’s not real, but one would imagine that, given the right character you could actually feel for such things, you kind of know it’s made up, but you get drawn in.*”

Whilst all of the users were sufficiently engaged to have social feelings about PACEO, many of them wondered whether such feelings would erode over time, until PACEO became an irritant. Participant E suggested: “*if PACEO’s ramblings get repetitive, if he fails to captivate me over time then he may become a little annoying.*” PACEO’s only hope to be continually used over a significant time would be to be highly useful and continually engaging.

6. Discussion

Giving PACEO the appearance of social role awareness appeared to make it more socially able and engaging. All participants attributed social ability to PACEO, yet, PACEO does not have social ability; it is neither an emotional nor an empathic entity. Rather, it is a fabrication, constructed through the fictional elaboration of its actual social context.

PACEO operates within the explicitly defined power structures of the organisation: its dynamic was one of good up relationships, varied peer relationships and poor down relationships. The agent appears to be successful at pleasing its superiors, ambivalent about its peers and negative towards its inferiors. As such, the agent is exhibiting the stereotypical profile of an upwardly aspirant member of the organization.

The apparent human-like behaviour exhibited by PACEO did not seem to give rise to the basis of a real social relationship; participants were not really prepared to consider PACEO as a social being. The agent software is not fooling the user into believing that the agent was alive and socially aware, rather something

more subtle is going on. The users seem to be aware that is just a program, but they seemed willing to become actors within the agent's fictive narrative.

The relationships that the users did create and describe were not the same as those that a user typically has with a tool (such as most software). Users seemed to anticipate a novel relationship with PACEO, which was neither functional nor social, with the user-agent relationship resulting in interactions and responses that were different to those experienced in human-human interactions and the majority of human computer interactions.

This behaviour seems more akin to a reader becoming involved in a plot or a player submerged within a gaming environment, than real human to human social interaction. However though we had not set out to create real social ability, our research does seem to be making the first tentative steps to creating an agent that is dynamic and engaging to its audience. As this evaluation occurred within a role-playing situation, it is not possible to predict that a similar situation will emerge in a live situation. However, this is something we intend to evaluate further.

Whilst there were limitations within the environment we used to test PACEO, the evaluation did enable us to rectify some usability and comprehension issues. It also allowed us to determine the feasibility of placing PACEO within the target organization and identified that once additional utterances had been developed that a real trial could occur.

The use of ethnography framed by Foucault's methodological approach was an invaluable tool in constructing the agent's narrative. Foucault was chosen because his tools are firstly sensitive to power and secondly focused on contemporary Western institutions. Primarily Foucault was used in the ethnographic phase of our research, but because he is concerned with how the social is delineated along binary divisions his method seems to allow a rudimentary transposition at the level of computer code.

It remains to be seen whether this approach will become a salient method for other agent research. However if social ability is indeed (as we believe) crucial for believable agent interaction, and if computers cannot deliver a true cultural entity, then the construction of cultural specificities as agent narrative seems at the present the only feasible way to achieve this.

References

- André, E. (1999). Editorial to Special Issue on Animated Interface Agents. *Applied Artificial Intelligence*, 13 (4/5), 341–342.
- Bates, J. (1994). The Role of Emotion in Believable Agents. *Communications of the ACM*, 37(7), 122–125.

- Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing Culture: The Poetics and Politics of Ethnography*. San Francisco, CA: University of California Press.
- Coates, J. (1994). The Language of the Professions: Discourse and Career. In A. Evettes (Ed.), *Women and Career: Themes and Issues in Advanced Industrial Societies* (pp. 195–230). London, UK: Longman.
- Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*: MIT Press.
- Foucault, M. (1977). *Discipline and punish – The birth of the prison*. New York: Penguin Books.
- Gutting, G. (1994). *The Cambridge Companion to Foucault*. Cambridge: Cambridge University Press.
- Hayes-Roth, B. (1998). Animate Characters. *Autonomous Agents and Multi-Agent Systems* 1(2), 195–230.
- Ho, W. C. & Watson, S. (2006) Autobiographic Knowledge for Believable Virtual Characters. In J. Gratch, M. Young, R. Aylett, D. Ballin, & P. Olivier (Eds.), *Intelligent Virtual Agents: 6th International Conference, IVA 2006* (pp. 383–394), LNCS 4133, Berlin/Heidelberg: Springer.
- Layton, R. (1997). *An introduction to theory in anthropology*. Cambridge, UK: Cambridge University Press.
- Mateas, M. (1997). An Oz-Centric Review of Interactive Drama and Believable Agents, 2003. www2.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/papers/CMU-CS-97-156.html
- McHoul, A. W. & Grace, W. (1997). *A Foucault Primer: Discourse, Power and the Subject*: New York University Press.
- Mitchell, M. D. (1999). *Governmentality: Power and Rule in Modern Society*: Corwin Press.
- Prada, R. & Paiva, A. (2005). Synthetic group dynamics in entertainment scenarios: creating believable interactions in groups of synthetic characters. In *Proc. of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE 2005)*, Valencia, Spain, June 15–17, 2005. ACM Press. Available at: <http://doi.acm.org/10.1145/1178477.1178585>
- Polaine, A. (2005). The flow principle in interactivity. In Y. Pisan (Ed.), *Proc. 2nd Australasian Conference on Interactive Entertainment*, ACM International Conference Proceeding Series; Vol. 123 (pp. 151–158). Sydney, Australia: Creativity and Cognition Studios Press.
- Prendinger, H. & Ishizuka, M. (2001). Let's talk! Socially intelligent agents for language conversation training. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 31(5), 465–471.
- Prendinger, H. & Ishizuka, M. (2002). Evolving Social Relationships with Animate Characters. In R. Aylett & L. Cañamero (Eds.), *AISB '02 Symposia: Animating Expressive Characters for Social Interactions* (pp. 73–78). SSAISB Press.
- Rabinow, P. (1984). *The Foucault Reader*. New York: Penguin Books.
- Rist, T. & Schmitt, M. (2002). Avatar Arena: An Attempt to Apply Socio-Physiological Concepts of Cognitive Consistency in Avatar-Avatar Negotiation Scenarios. In R. Aylett & L. Cañamero (Eds.), *AISB '02 Symposia: Animating Expressive Characters for Social Interactions* (pp. 79–84). SSAISB Press.
- Spiro, M. E. (1996). Postmodernist Anthropology, Subjectivity, and Science. A Modernist Critique. In E. James & G. Marcus (Eds.), *Comparative Studies in Society and History* (pp. 759–780). University of California Press.

CHAPTER 3

What's in a robot's smile?

The many meanings of positive facial display

Marianne LaFrance

1. Introduction

Facial expressions bespeak consciousness. A scowl or smile on a human face allows a perceiver to infer something about another person's inner experience. Scowls and smiles also allow perceivers to imagine what the other is likely to do as well as plan how one might or might not engage the other. In short, humans take facial expressions very seriously because they provide a theory about another's person mind. Can facial expressions from a robot be taken as seriously? Can robots learn to take human facial expressions seriously? Apparently, the answer is a tentative yes. At MIT, when people confront the robot Kismet, with its moveable red lips and fuzzy brows and big ears, they act as though it has feelings, that social interaction with it is possible (Whynott 1999).

The possibility of animating meaningful expressive behavior in robots is a relatively new question with a fairly substantial past. From research in social perception, we know that people respond to mobile inanimate objects as though they have intentions (Michotte 1963). From early work in artificial intelligence, we know that people believe a computer program has intelligence if it simulates the mode of a non-directive psychotherapist. Specifically, the computer program (called ELIZA) responded to statements by people with questions. If someone keyed in the phrase, "I'm feeling sad," the response from the computer was, "Why are you feeling sad?" Such simple responses led users to believe that the computer possessed special insight into what they were feeling (Dreyfus & Dreyfus 1986). What remains to be understood is how people interpret human-like emotive visual displays when inanimate characters display them. And how might people respond if robots appear not only to respond to verbal input from them but also respond to what their faces show. This chapter takes one particular facial display, namely the smile, and discusses what robots will have to "understand" about smiling so that both humans and robots can take each other's smiles seriously.

Among humans, smiles are distinctive, complex, and nearly indispensable. People have been known to go to considerable lengths to get someone to produce a smile and then feel positively euphoric when they have been successful in doing so. Smiling is easily detected and can be perceived even when it surfaces for very brief periods and from considerable distances (Ekman & Hager 1979). Darwin suggested that because smiles are so different in appearance from negative expressions, they became associated with happiness (Darwin 1872/1965). Nonetheless, human smiles can themselves take a variety of forms, exhibit an array of intensities and durations, and can be found across a broad set of circumstances (LaFrance *et al.* 2003). Among humans, although smiling is sometimes spontaneous, it can also be both employed deliberately according to rules and rituals about how one should smile and when it is legitimate to do so. For example, smiling is prohibited in some situations and it is obligatory in others (Wierbicka 1994). In short, human smiling in all its various guises is probably indispensable for the creation and maintenance of social relationships (LaFrance & Hecht 1999).

Creators of social robots have taken note of the importance of smiling (Breazeal 1998; Billard & Dautenhahn 1997). For example, Breazeal's robot Kismet has been designed to show emotive expressions analogous to happiness (along with other emotions like anger, fear, disgust, excitement, sadness, and surprise). Specifically, Kismet is programmed to respond to external visual input with changes in its facial expressions. Such emotive displays apparently demonstrate to human "care-givers" that it possesses the capacity to communicate with them. While initial tests of these capabilities show promise, nonetheless, it is also clear that expressive communication between robots and people will only really work to the extent to which such displays are credible to humans. Breazeal (2000) has phrased it this way, "For a robot, an important function of the motivation system is to regulate behavior selection so that the observable behavior appears coherent, appropriately persistent, and relevant given the internal state of the robot and the external state of the environment."

In what follows, I draw on recent research in the social psychology of the smile to show that the task of creating credible facial expressions in robots will be a formidable one. Equally daunting will be the task of programming social robots to recognize differences in smiles on people's faces, that is, to distinguish their various forms, configurations, intensities and meanings. Specifically, I describe various aspects of human smiles that need to be taken into account if robots' smiling is to be taken seriously. To be sure, human smiles reflect and generate positive affect (Ekman & Friesen 1982) but that is not all they are or do. In fact, there is good reason to believe that human facial expression in general and smiles in particular serve many psychological and social purposes beyond the simple expression of positive affect. In what follows, I describe some of the various forms

smiles take, some of the range of functions that human smiles serve, and some of the range of social meanings they convey.

2. Smiles: Indicators or messages?

The meaning of smiling seems obvious – people smile when they are happy or amused or delighted (Ekman & Friesen 1982; Ekman *et al.* 1990). So too goes the logic for other facial expressions: people frown when they are displeased and sneer when they are disgusted and grit their teeth when they are angry. The idea goes back to Aristotle (n.d. 1913) who noted: “There are characteristic facial expressions which are observed to accompany anger, fear, erotic excitement, and all the other passions.” A contemporary philosopher espoused a similar idea. Merleau-Ponty (1961/1964: 52–53) wrote: “Anger, shame, hate, and love are not psychic facts hidden at the bottom of another’s consciousness; they...exist on this face or in those gestures, not hidden behind them”. These emotion-based models of facial expression, referred to sometimes as readout models (Buck 1991) or facial expression programs (Russell & Fernández-Dols 1997) share the core assumption that the face provides an external read out of internal emotional or affective states.

Emotionally incongruent smiles sometimes *appear* indistinguishable from smiles associated with happiness and sometimes they appear on the face as a blend in which a smile is only one part of a more complex facial configuration. In the former case, a smile is not a direct indicator about what a person is really feeling at the moment. Instead the smiles are volitional (though not necessarily conscious) and have been displayed to be communicative. The message may be “this is how I want you to see me (even if I don’t really feel this way)” or “I am displaying this expression because it is appropriate to do so in this particular social situation”. Such smiles may look similar to smiles that actually indicate underlying positive emotion. Yet it is important to remember that they may not actually do so. And accurate interpretation of the meaning of smile often requires that perceivers know the difference.

Close inspection may reveal that the expression is a deliberate message rather than a spontaneous indicator. For example, if a smile display is asymmetrical, that is, stronger on one side of the face than the other side, then it is likely that the expression is in some sense deliberate (Frank *et al.* 1993). Or if the onset or offset of the smile is too abrupt or if it stays on the face too long, then too there is the suggestion of a premeditated display. Because spontaneous smiles (indicators) and volitional smiles (messages) sometimes look the same should not be construed that they mean the same thing or that they have the same effect on self or others. In sum, some smiles may reflect an underlying happy emotion but other

smiles, which may look very much the same, are better described as being messages about positive intentions rather than actually indicating positive feelings. In the former case, a smile is an *indicator*; at other times, it is a message directed by a sender to a receiver sent in the interests of the sender.

The distinction between facial expression as indicator or message has sometimes been characterized as the distinction between emotion-based expression and social-based display. In arguing for the distinction, researchers have attempted to show that smiling among adults is more often socially-cued than emotion-cued. In other words, they contend that facial displays have no necessary relation to emotions but have everything to do with being seen in a particular way. For example, Fridlund and his colleagues have found that people smile more when others are present than they do when alone even when the level of felt positive emotion is the same. They have also conducted experiments where people are compared on how much smiling they exhibit when they are alone but believe that a friend is nearby than when they do not believe that to be the case. They find that people smile more even when they imagine being in the presence of someone they know (Fridlund *et al.* 1990).

Kraut and Johnston (1979) have explored a similar idea, namely that smiling is not an automatic readout of happy emotion but is, rather, a social message meant to be seen. Specifically, they took careful note of when and at what or whom people smile when they are bowling. They noted that bowlers who had just succeeded in getting a strike smiled more when they turned to face their friends than immediately following the successful throw. In short, a good deal of smiling is done with a purpose, that is, to establish and effect certain kinds of social connections rather than as an automatic manifestation of positive feeling.

2.1 Smile types

Sometimes one is pleased – the feeling is an unequivocally happy one. So too, some smiles are unambiguous signs of positive emotion as shown by Figure 1.

However, among emotion researchers there is considerable debate about whether humans should best be described as possessing a small set (six) of basic emotions that are associated with a respective set of particular and unambiguous facial displays. For example, a number of investigators have argued that emotion is better conceptualized as a dynamic process in which the connection between emotion and expression is malleable and context dependent (Russell 2003; Turner & Ortony 1992). In other words, human beings do not possess a single set of basic pure emotions that are associated with a single set of unambiguous facial displays. Rather, there are families of emotion and families of associated facial expressions



Figure 1. Example of Duchenne (real) smile

(Fogel *et al.* 1992). Smiles appear to be a large family of facial expressions. In fact, Ekman and Friesen (1982) contend that smiles may be the most misunderstood of all facial expressions precisely because there are so many kinds.

2.2 Duchenne and non-Duchenne smiles

Ekman and Friesen (1982) initially distinguished between two broad smile types, namely the *felt* smile and the *false* smile. In subsequent work, these were renamed the *Duchenne* and *non-Duchenne* smiles respectively. In Duchenne smiles, as shown in Figure 1, lip corner contraction is accompanied by contraction of the muscles surrounding the eyes, which causes the cheeks to rise and crows feet wrinkles to appear at the corners of the eyes.

For adults, Duchenne smiling has been empirically linked to the experience of positive emotion. For example, Duchenne smiling is more likely to occur when people are watching pleasant as compared to unpleasant films. Duchenne smiling compared with other smiles is more strongly correlated with people reporting feelings of amusement, happiness, excitement, and interest (Ekman *et al.* 1980). Observers also rate Duchenne smiles as more positive compared to other smiles, which is why they were initially labeled as felt smiles (Ekman *et al.* 1990).

In contrast, non-Duchenne smiles are recognized by a single facial action, namely the contraction of the lip corners with no discernible facial movement around the eyes. The term “non-Duchenne smile” replaced the “false” smile



Figure 2. Moderate non-Duchenne smile

designation to allow for the fact that this type of expression is not necessarily deceptive or unreal. It is true that non-Duchenne smiles are not reliably associated with the experience of amusement or happiness in the person who shows them but it is also true that these smiles serve a number of important social functions. This type of social smile also appears to be much more frequent than Duchenne smiles. We examined the effect of different types of smiles shown by a target person believed to have committed some misconduct. We found that persons who showed Duchenne and non-Duchenne smiles were judged more leniently than persons who did not smile. However, we also found that non-Duchenne (false) smiles actually generated more benefit of the doubt for the transgressor than Duchenne (felt) smiles. Analyses showed that those who displayed the non-Duchenne smiles were rated as more trustworthy than those who showed Duchenne smiles. We conjectured that this was due to the possibility that a truly happy expression in this context was perceived as somewhat inappropriate. Rather the non-Duchenne smile was acknowledged by observers to be a socially constructed display – a display that recognized that others were sitting in judgment of them (LaFrance & Hecht 1995).

Other data show that the distinction between Duchenne and non-Duchenne smiling is a socially meaningful one. Findings show that these two smile types often appear in different social contexts and correlate with different patterns of brain electrical activation (Ekman *et al.* 1990). Developmental psychologists have found that very young infants show both Duchenne and non-Duchenne smiles and these different smile types are associated with distinctly different contexts (Dickson *et al.*

1997). For example, Duchenne smiles by 10-month old infants have been found to be associated with the approach of their mothers whereas non-Duchenne smiles are associated with the approach of strangers (Fox & Davidson 1988). Fogel and his colleagues contend that non-Duchenne smiles in infants represent a particular kind of enjoyment, namely the enjoyment that comes with being *ready* to engage in play. Readiness to engage in play appears to be a different emotional state accompanied by a different facial expression than the enjoyment that comes with actually engaging in play (Fogel *et al.* 2000, p. 513).

Studies of infants have also discerned two additional smile types; namely the *play* smile and the *duplay* smile (Dickson *et al.* 1997). In play-smiles, the baby's jaw drops along with the lip corners rise. A duplay smile adds a cheek raise to the lip corner rise and the jaw drop. The research finds that these different smiles are associated with different social contexts. For infants and toddlers, the play-smile is more common while playing with an attentive partner, whereas the non-Duchenne smile occurs more frequently during toy-centered play. Duplay smiles are more likely to occur during highly arousing physical play (often with fathers) and during book reading (often with mothers). During physical play another nonverbal element is added to the mix: accompanying the mouth opening there is often a discernible inhalation (Dickson *et al.* 1997). In fact, both play and duplay smiles are often accompanied by vocalizations such as laughs, squeals and yells.

Smiles by 12-month olds can be differentiated on the basis of other accompanying behaviors by infants and by people they are interacting with. For example, duplay smiling is more likely to occur while the infant is looking at the mother and when the mother is smiling, such as might occur when parent and child are playing a game of peek-a-boo. That particular configuration appears to indicate enjoyment in the emotion build-up. In sum, different co-occurring interpersonal actions and distinguishable facial configurations appear to be associated with different emotional states and interpersonal dynamics.

The meaning of different smile configurations also changes as children get older. When pre-schoolers are engaged in an achievement-oriented game, smiling seems to be a good indicator of how much joy or pride the child takes in doing well. But when the game no longer occupies the child's complete attention and when a familiar person is present, then the smiling seems to have a more communicative (symbolic) function, such as commenting on the game or belittling failure at it. In other words, children become increasingly able and willing to control their facial behaviors with the result that the connection between the emotion and expression becomes less directly coupled.

2.3 Smile-blends

In adults, research has shown that the familiar lip corner raise recognized as the prototypical smile can co-occur with other visible changes on the face, changes associated with emotion states other than positive affect. A lip corner raise combined with a raised forehead often accompanies the feeling of surprise. When elements from different emotional states are simultaneously present on the face, the result is called an expressive blend. A case in point is the *miserable smile blend* shown in Figure 3 where a retraction of the lip corners co-occurs with two other facial actions, namely lowered eyebrows and a middle-forehead rise. The resulting blend shows distress in the eyes and a smile on the mouth. According to Ekman (1985), the miserable smile acknowledges that while the sender is experiencing distress, he or she is simultaneously declaring to others that one has the ability to handle it.

Lip corner retraction, typically associated with positive feeling, can also combine with facial changes signaling anger (lowered brows, open-mouth and clenched teeth). The angry smile such as shown in Figure 4 might transpire when the prevailing emotion involves taking pleasure in somebody else's misfortune, the sentiment sometimes referred to as *schadenfreude*.

Embarrassed smiles (see Figure 5) are characterized by the co-occurrence of gaze aversion, downward head movement and by indications that the smile is itself being controlled (Keltner 1995).

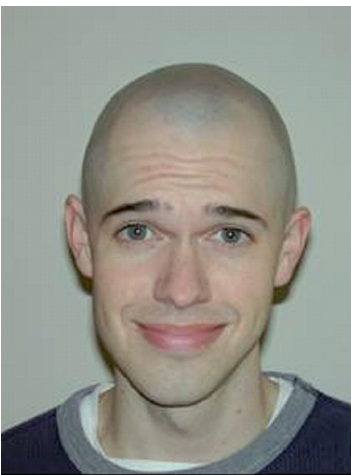


Figure 3. Miserable smile

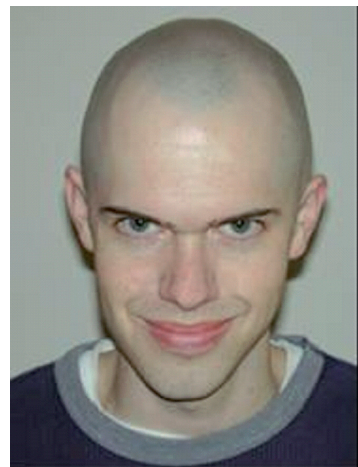


Figure 4. Angry smile

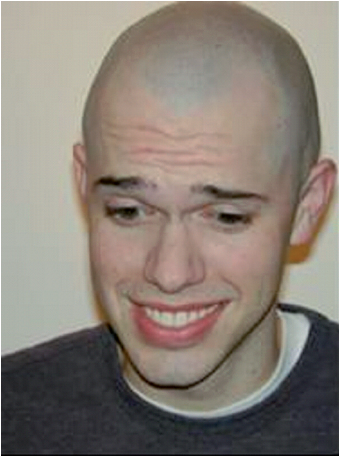


Figure 5. Embarrassed smile

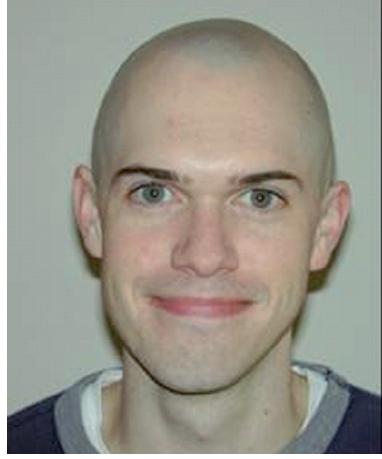


Figure 6. Dimpler

According to Keltner, embarrassment smiles do not merely serve as the outward manifestation (i.e., sign) of feeling mortified but they signal submissive behavior. Empirical evidence shows that displays of embarrassment serve the social function of appeasing more dominant individuals by evoking affiliation and forgiveness.

2.4 Smile look-alikes

There are also facial actions involving the mouth that have some similarity to those associated with smiling but nonetheless involve different muscle contractions. One of these has been called the “dimpler” (Ekman & Friesen 1978), depicted in Figure 6. As Figure 6 shows, the dimpler involves contraction of the buccinator muscle, which tightens the lip corners often causing a dimple to form near the lip corners. The important issue to take note of here is that smiles appear in a variety of forms and findings strongly indicate that subtle yet discernible changes in what looks like a smile are associated with meaningful different emotional and social circumstances.

3. Smiling and social context

The discussion thus far has centered on the idea that there is not one kind of smile, which is unequivocally indicative of positive emotion. There are different kinds

of smiles associated with various emotional states and various social imports. But another part of the complexity in understanding the nature of human smiling derives from recognizing that smiling is intimately and intricately tied to social context. In the next section I address three important social context dimensions.

3.1 Smiling and social obligation

It appears that smiling is prohibited in some social contexts and required in others even if, or perhaps especially if, people are not feeling particularly happy. For example, anthropologists and sociologists have documented numerous instances where people smile because the social situation or role or occupation requires them to do so. For example, greetings are frequently associated with smiling, so much so that the absence of smiling in brief encounters among mere acquaintances is enough to signal discourtesy or displeasure (Eibl-Eibesfeldt 1989).

Within the United States, Hochschild (1983) noted that many workers are required to smile as part of their jobs. For instance, airline flight attendants must smile and smile well. Thus a flight attendant is trained to “really work on her smiles” and is expected to “manage her heart” in such a way as to create a smile that will both seem as “spontaneous and sincere” (Hochschild 1983, p. 105).

There is also evidence that smiling and expressions of interest are expected in social encounters where another is telling a story about a happy occurrence. We examined this idea by asking males and females to imagine smiling or not in a number of different social contexts. For example, they were instructed to imagine that they did or did not smile in response to someone reporting some good news. We predicted that both sexes but especially women would anticipate more negative repercussions when they did not smile (LaFrance 1997), and that is what we found. People felt significantly less comfortable and less appropriate when they did not smile in response to someone else’s good news. Women in particular believed that another’s impression of them would change more if they did not smile. Other studies find too that people expect fewer rewards and higher costs when their responses to others are insufficiently positive (Stoppard & Gunn-Gruchy 1993).

In still other social contexts, smiling is displayed because smiling elicits acquiescence from others. Smiles are linked with attempts to persuade someone to think or feel a particular way (Burgener *et al.* 1992). For example, smiling is a common behavior among politicians (Masters *et al.* 1985). Smiles have also been shown to ward off others’ displeasure (Elman *et al.* 1977; Goldenthall *et al.* 1981).

3.2 Smiles as cultural displays

Over thirty years ago, Ekman and Friesen (1969) suggested that people do not always show what they feel. What gets expressed facially is guided by display rules that dictate to whom and in what contexts it is appropriate or inappropriate to show various expressions. In particular, cultures vary substantially in the degree to which they direct members about how they should show their faces in public. Wierzbicka (1994) has observed for instance that cheerfulness is mandatory in many cultures. In other groups, the same amount of smiling might be seen as infantile, condescending or inappropriate. Indeed there is a growing literature indicating that expressive displays vary substantially across national, ethnic and regional boundaries (Mesquita 2001). For example, a recent study compared the amount of smiling shown by Chinese American and European American couples as they discussed problems in their relationships. Chinese American couples were found to display less positive emotional expression than European-American couples but no more negative expressions (Tsai *et al.* 1999). Hall and Oram (this volume) also draw attention to the fact that embodied agents operate within a culture.

3.3 Smiles as gender displays

Will emotionally expressive robots have a sex? Although Kismet is described as gender neutral, it is not likely to stay that way for long. And when it gets described as “he” or “she”, a whole new set of expectations about appropriate levels and types of smiling will come to the fore.

One of the most consistent empirical findings in the current literature on smiling is that women smile more than men do at least in Western European cultures (LaFrance *et al.* 2003). However the reasons for this greater expressivity are many. To begin with, females are believed to be more emotional than men are and socialization practices tend to encourage more attention to emotional matters among girls than among boys. Other speculations have to do with the finding that many social roles and occupations are segregated by gender. Females are more often than males assigned to tasks and jobs that have emotional expressivity as a core requirement.

4. Implications for the design of affective displays in robots

One of the most interesting things about human smiling is the range of forms it can take, the range of contexts in which it can be found, and the range of social

meanings that can be ascribed to it. To be sure, smiling is associated with positive emotion, but smiling can also be associated with its opposite. Smiling is often encouraged and appreciated but there are times when it is discouraged and suppressed. Smiles are used to influence what will happen as well as simple expressions of what has happened. In sum, if smiles on robots are to be taken seriously, researchers will need to incorporate research findings on how, when, and why people smile.

The hope in this chapter has been to prompt researchers to continue to examine the implications of endowing expressive characters with human-like expressive abilities. To this end, I offer a few recommendations.

First, developers of expressive robots would do well to consider alternative frameworks for understanding expressive displays. At present, the preferred model among computer scientists is an emotion read-out model. According to such models, emotion precedes expression and each emotion is associated with a distinctive expressive display. I suggest that much could be gained by conceptualizing expressive display as a message system and not merely as an indicator of a particular underlying emotion. In developing expressive capabilities of robots, the question should change from what emotions do we want them to display to what messages do we want them to send and what actions do we want it to induce. Receivers of expressive displays use them to ascertain information about the sender's current situation, future actions, prevailing attitudes and cognitive states not merely its emotional state (Russell *et al.* 2003). Clearly research is needed on what information human recipients extract from robots' expressive displays.

My second recommendation follows quite closely from the first, namely the need to study the meanings attributed to and consequences evoked by robots' expressive displays by different people in different contexts and over time. Given advances in the study of the meaning of nonverbal expression, it is very important to move beyond the demonstration that smiling expressions elicit approach and frowning ones, withdrawal. We still know too little about the consequences of expressive displays both short term and long term. The immediate reaction to a smiling robot may be different from later reactions when conscious cognitive appraisal becomes more likely.

Third, developers of expressive capabilities in robots would do well to move beyond relatively static conceptions of facial displays. As research on smiles has shown, timing and composition have everything to do with attributed meaning. These are not meaning add-ons but essential features of the messages themselves. Finally, it will be important for researchers in this area to consider which cultural, social, and gender practices expressive characters will adopt. Ruttkay and colleagues (this volume) also note that for embodied conversational agents to be credible they will need to take account of culture, gender, age, personality as well

as physical state and speaker mood. Were these suggestions be put into effect, we would know both more about the possibility of credible expressive capabilities of robots as well as a great deal more about expressivity in humans.

References

- Billard, A. & Dautenhahn, K. (1997). Grounding Communication in Situated, Social Robots. Report No. UMCS-97-9-1. Manchester, England: University of Manchester.
- Buck, R. (1991). Social Factors In Facial Display and Communication: A Reply to Chovil and Others. *Journal of Nonverbal Behavior*, 15(3): 155–161.
- Breazeal, C. (1998). A Motivational System for Regulating Human-Robot Interaction. *Proceedings of the Fifteenth National Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press, 54–61.
- Breazeal, C. (2000). Infant-like Social Interactions between a Robot and a Human Caregiver. *Adaptive Behavior* 8: 9–75.
- Bugener, S. C., Jirovec, M., Murrell, L., & Barton, D. (1992). Caregiver and environmental variables related to difficult behaviors in institutionalized demented elderly persons. *Journal of Gerontology* 47: 242–249.
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. New York: Appleton 1965.
- Dickson, K. L., Fogel, A., & Messinger, D. (1997). The Development of Emotion from A Social Process View. In M. Mascolo and S. Griffen (Eds.), *What develops in emotional development* (pp. 253–273). New York: Plenum Press.
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: The Free Press.
- Edelmann, R. J., Asendorpf, J., Contarello, A., & Zammuner, V. (1989). Self-Reported Expression of Embarrassment in Five European Cultures. *Journal of Cross Cultural Psychology*, 20: 357–371.
- Eibl-Eibesfeldt, I. (1989). *Human Ethology*. New York: Aldine de Gruyter.
- Ekman, P. (1985). *Telling Lies*. New York: Norton.
- Ekman, P., Davidson, R. J., & Friesen, W. (1990). The Duchenne Smile: Emotional Expression and Brain Physiology II. *Journal of Personality and Social Psychology*, 58: 342–353.
- Ekman, P. & Friesen, W. V. (1969). The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica*, 1: 49–98.
- Ekman, P. & Friesen, W. V. (1978). *The Facial Action Coding System (FACS): A Technique for the Measurement of Facial Action*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P. & Friesen, W. V. (1982). Felt, False, And Miserable Smiles. *Journal of Nonverbal Behavior*, 6: 238–252.
- Ekman, P., & Hager, J.C. (1979). Long Distance Transmission of Facial Affect Signals. *Ethology and Sociobiology*, 1: 77–82.
- Elman, D., Schulte, D. C., & Bukoff, A. (1977). Effects of Facial Expression and Stare Duration on Walking Speed: Two Field Experiments. *Environmental Psychology and Nonverbal Behavior*, 2: 93–99.

- Fogel, A., Nelson-Goens, G. C., & Hsu, H.-C. (2000). Do Different Infant Smiles Reflect Different Positive Emotions? *Social Development*, 9: 497–520.
- Fogel, A., Nwokah, E., Dedo, J., Messinger, D., Dicson, K. L., Matusove, E., & Holt, S. A. (1992). Social Process Theory of Emotion: A Dynamic Systems Approach. *Social Development*, 1: 122–142.
- Fox, N. A. & Davidson, R. J. (1988). Patterns Of Brain Electrical Activity During the Expression of Discrete Emotions In 10-Month-Old Infants. *Developmental Psychology*, 24: 230–236.
- Frank, M., Ekman, P., & Friesen, W. V. (1993). Behavioral Markers and Recognizability of the Smile of Enjoyment. *Journal of Personality and Social Psychology*, 64: 83–93.
- Fridlund, A. J. (1991). Sociality of Solitary Smiling: Potentiation by an Implicit Audience. *Journal of Personality and Social Psychology*, 60: 229–240.
- Fridlund, A. J., Sabini, J. P., Hedlund, L. E., Schaut, J. A., Shenker, J. I., & Knauer, M. J. (1990). Audience Effects on Solitary Faces During Imagery: Displaying To the People in Your Head. *Journal of Nonverbal Behavior*, 142: 113–137.
- Hecht, M. A. & LaFrance, M. (1998). License or Obligation to Smile: Power, Sex and Smiling. *Personality and Social Psychology Bulletin*, 24: 1326–1336.
- Hochschild, A. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley, CA: University of California Press.
- Keltner, D. (1995). The Signs of Appeasement: Evidence for the Distinct Displays of Embarrassment, Amusement, And Shame. *Journal of Personality and Social Psychology*, 68: 441–454.
- Kraut, R. E. & Johnston, R. E. (1979). Social and Emotional Messages Of Smiling: An Ethological Approach. *Journal of Personality and Social Psychology*, 37: 1539–1553.
- LaFrance, M. (1997). Pressure To Be Pleasant: Effects Of Sex And Power On Reactions To Not Smiling. *Revue Internationale de Psychologie Sociale/International Review of Social Psychology*, 2: 95–108.
- LaFrance, M. & Hecht, M. A. (1999). Obligated To Smile: The Effect Of Power And Gender On Facial Expression. In P. Philippot, R. S. Feldman, & E. J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 45–70) Cambridge: Cambridge University Press.
- LaFrance, M. & Hecht, M. (1995). Why Smiles Generate Leniency. *Personality and Social Psychology Bulletin*, 21: 207–214.
- LaFrance, M., Hecht, M., & Paluck, E. (2003). The Contingent Smile: A Meta-Analysis of Sex Differences in Smiling. *Psychological Bulletin*, 129(2): 305–334.
- Masters, R. D., Sullivan, D. G., Lanzetta, J. T., McHugo, G. J., & Englis, R. (1986). The facial displays of leaders: Toward an ethology of human politics. *Journal of Social and Biological Structures*, 19: 319–343.
- Mesquita, B. (2001). Culture and Emotion: Different Approaches to the Question. In T. J. Mayne & G. A. Bonanno (Eds.), *Emotions: Current Issues and Future Directions* (pp. 214–250). New York: Guilford Press.
- Michotte, A. (1963). *The perception of causality*. New York: Basic Books.
- Ochanomizu, U. (1991). Representation Forming In Kusyo Behavior. *Japanese Journal of Developmental Psychology*, 2: 25–31.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110: 145–172.
- Russell, J. A., Bachorowski, J., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54: 329–349.

- Russell, J. A. & Fernández-Dols, J. M. (1997). What Does a Facial Expression Mean? In J. A. Russell & J. M. Fernández-Dols (Eds.), *The Psychology of Facial Expression* (pp. 3–30). Cambridge, England: Cambridge University Press.
- Stoppard, J. M. & Gunn-Gruchy, C. D. (1993). Gender, context and expression of positive emotion. *Personality and Social Psychology Bulletin*, 19: 143–150.
- Tsai, J. L., Levenson, R. W., & McCoy, K. (1999). Are Chinese Americans less emotional than European Americans? Culture, Context, and Components of Emotion. Unpublished manuscript. University of Minnesota, Minneapolis.
- Turner, T. J. & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review*, 99: 566–571.
- Wierzbicka, A. (1994). Emotion, Language And Cultural Scripts. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 133–196). Washington, DC: American Psychological Association.
- Whynott, D. (1999). It Smiles, It Frowns – It's A Robot! *Science World*, 56: 8–10.

CHAPTER 4

Facial expressions in social interactions

Beyond basic emotions

Susanne Kaiser and Thomas Wehrle

1. Introduction

In recent years, the importance of emotions and emotion expression in human interactions has been widely accepted. Studying the functions of facial expressions is not only of interest in the domain of psychology and other social sciences but also in the domain of computer science and computational intelligence. An increasing number of research groups are developing computer interfaces with synthetic facial displays (e.g., Ball & Breese 2000; Elliot 1997; Takeuchi & Naito 1995; Pelachaud 2005). These researchers attempt to use facial displays as a new modality that should make the interaction more efficient, while lessening the cognitive load on the user. In addition, several researchers point out that automatic interpretation of gestures and facial expressions would also improve human-machine interaction (e.g., Essa & Pentland 1997; Lien, Kanade, Zlochow, Cohn & Li 1998).

The relatively new field of “affective computing” (Picard 1997), located between psychology, engineering, and natural sciences, demonstrates the promise for interdisciplinary collaboration in the domains of emotional facial expression synthesis and automatic expression recognition. Although the attempt to implement specific models of the emotion process is not new (for reviews see Pfeifer 1988; Wehrle 1994a; Wehrle & Scherer 1995; Wehrle & Scherer 2001), the availability of powerful techniques in artificial intelligence and the increasing focus on enhanced user interfaces render this approach particularly promising nowadays. This book, which originated in an interdisciplinary workshop on animating expressive characters for social interactions, also demonstrates this progress (see also Cañamero 2001, 2005; Hudlicka & Cañamero 2004; Paiva 2000; Paiva *et al.* 2007; Pelachaud & Cañamero 2006; Tao *et al.* 2005). Within this field, an important distinction has to be made between artificial emotions on the one hand, and theory modeling, on the other. Whereas the first domain is concerned with the application of emotion theories to user interfaces and to the design of autonomous emotional agents, theory modeling serves the purpose of improving our

understanding of the phenomenon of emotion in humans (for more details see Wehrle 2001; Wehrle & Scherer 2001). Although our research interests are more linked to the latter purpose, we think that an interdisciplinary approach is useful and also necessary in order to develop efficient “emotional interfaces”. This chapter aims to illustrate how an appraisal-based approach to the understanding of the relation between emotion and facial expressions might be instrumental to these different domains and their possible applications, namely (a) facial expression synthesis (animated intelligent agents), (b) automatic expression recognition (decoding the emotional state of the user), and (c) computational modeling of emotions (theory modeling and artificial emotions).

2. Cognition-emotion interaction: Emotional problem solving

We are using human-computer interactions and interactive computer games in order to study the ongoing dynamic cognitive and emotional processes including situated behavior and verbal and nonverbal expression. The experimental setting allows us to study the dynamics of emotional episodes in a more interactive manner than it is usually done in classical experimental settings.

One reason why traditional emotion induction techniques often trigger only rather weak and not clearly defined emotional states might be that subjects are not really involved and/or that they have to verbalize their emotional state instead of reacting to an emotion eliciting event or situation. This is diametrically opposed to one of the most adaptive functions of our emotionality – to guide our behavior in situations that are important for our goals and needs and that require immediate cognitive and behavioral responses. In real life situations, we are usually not capable of developing “the perfect solution” to a problem; related to this idea is the notion of “bounded rationality”, introduced by Herbert Simon about forty years ago and amply developed in (Gigerenzer & Selten 2001). We cannot consider every aspect of imaginable decisions or reactions, such as all possible short- and long-term consequences. Yet, in many cases, emotions help us to find “good” solutions. We refer to this as *emotional problem solving*. Emotional problem solving is not a “yes-or-no” decision making task but a process that unfolds in emotional episodes. Decisions are adapted and changed according to the dynamic changes in the external environment and according to changes caused by internal processes, concerning for instance memory, motives, and values. Following cognitive emotion theories, we can describe an emotional episode as a process of primary appraisal (the subjectively estimated significance of an event for one’s well being), secondary appraisal (the subjectively estimated ability to cope with the consequences of an event), coping, and reappraisal in a transactional inter-

action (Lazarus 1966). As suggested by Leventhal and Scherer (1987), these appraisal processes can occur at different levels of processing and are often very fast and automatic. Appraisal processes occurring on the sensory-motor or schematic level are rarely or only with great difficulty accessible through verbalization. One reason for analyzing facial expressions in emotional interactions is the hope that these processes might be accessible or indicated by facial expressions. Another reason for analyzing facial expressions in experimental emotion research is that these are naturally accompanying an emotional episode, whereas asking subjects about their feelings interrupts and changes the ongoing process.

3. The multi-functionality of facial behavior

Facial expressions can have different functions and meanings (see Ekman & Friesen 1969; Fridlund 1994; LaFrance, this volume; Russell & Fernández-Dols 1997). For example, a smile or a frown can be:

- A speech-regulation signal (*regulator*): a listener response (back-channel signal), telling the speaker that he can go on talking and that his words have been understood.
- A speech-related signal (*illustrator*): a speaker can raise his eyebrows in order to lay particular emphasis on his or her argumentation. The facial signals can also modify or even contradict the verbal messages, e.g., a smile that indicates that what is being said is not meant to be taken seriously.
- A means for signaling relationship: installing, maintaining, or aborting a relationship, e.g., when a couple is discussing a controversial topic, a smile can indicate that although they disagree on the topic, the relationship itself is not at stake.
- An indicator of cognitive processes: for example, frowning often occurs when somebody does some hard thinking while giving attention to a problem, or when a difficulty is encountered in a task.
- An indicator for an emotion (*affect display*): a person smiles because he or she is happy. Besides, affect displays that occur during an interaction can refer to the interaction partner (e.g., becoming angry with the other) but they can also refer to other persons or topics that the interaction partners are talking about (e.g., sharing their anger about something).

When considering spontaneous interactions, it is very difficult to identify whether a facial expression is an indicator of an emotion (affect display) or whether it is a communicative signal. To make things even more complicated, a facial expression can have several meanings at the same time; a frown, for instance, can

simultaneously indicate: (a) that the listener does not understand what the speaker is talking about (cognitive difficulty); (b) a listener response (communicative function) signaling that the speaker has to explain his argument more appropriately; and (c) that the listener is becoming more and more angry (emotional) about this difficulty in understanding him, about the content, or about the way this interaction develops.

Considering the complex and multiple functions of facial expressions, research paradigms for studying emotions and facial expressions should fulfill the following requirements:

1. To measure facial expressions objectively and on a micro-analytic level. The Facial Action Coding System (FACS; Ekman & Friesen 1978) lends itself to this purpose. FACS allows the reliable coding of any facial action in terms of the smallest visible unit of muscular activity (Action Units), each referred to by a numerical code. As a consequence, coding is independent of prior assumptions about prototypical emotion expressions. Using FACS we can test different hypotheses about linking facial expression to emotions.
2. The current, concrete meaning of a facial expression can only be interpreted within the whole temporal and situational context. In everyday interactions, we know the context and we can use all information that is available to interpret the facial expression of another person. Therefore, facial expressions and emotions should be studied in an interactive context.

We refer to coding procedures that fulfill both requirements as *situated coding procedures*. Situated coding procedures are anatomically based, objective procedures that are part of a given experimental setting. Besides facial expression data, the emotion eliciting experimental context as well as additional behavioral records (multimodal measurements) is systematically coded and this data is used for the interpretation of the facial behavior (for more details see Kaiser & Wehrle 2000).

4. Contemporary emotion theories

Most contemporary emotion theorists (for example Ekman 1992; Frijda 1986; Izard 1991; Scherer 1984) consider emotion as a phylogenetically continuous mechanism for flexible adaptation, serving the dual purpose of rapid preparation of appropriate responses to events and of providing opportunities for re-evaluation and communication of intent in the interest of response optimization. Following Darwin's (1876/1965) view of expressions as rudiments of adaptive behavior which have acquired important signaling characteristics, a functional approach to the study of emotion presumes that motor expressions are at the same time

reliable external manifestations of internal affective arousal and social signals in the service of interpersonal affect regulation (see also Kaiser & Scherer 1998; Bänninger-Huber & Widmer 1996; Roseman & Kaiser 2001; Kaiser 2002).

There is a long tradition in emotion psychology of examining facial expressions as an observable indicator of unobservable emotional processes. Among emotion theories proposing implicit or explicit predictions for emotion-specific facial expression patterns, two positions can be distinguished. The first approach is situated in the tradition of *discrete emotion theories* and is represented by Ekman, Izard, and their respective collaborators (e.g., Ekman 1992; Izard 1991). The second approach has been suggested in the context of *appraisal theories of emotion* (e.g., Kaiser & Wehrle 2001; Roseman, Wiest & Swartz 1994; Scherer 1992; Smith & Ellsworth 1985; Wehrle, Kaiser, Schmidt & Scherer 2000).

4.1 Discrete emotion theory and facial expression: Basic emotions

Discrete emotion theorists have studied facial expression as the “via regia” to emotions for many years. Most of their research concerning the universality of the so-called *basic emotions* is based on studies about facial expressions. These theories claim that there are only a limited number of fundamental or basic emotions and that, for each of them, there exists a prototypical, innate, and universal expression pattern. In this tradition, a process of blending or mixing the basic expression patterns explains the variability of emotion expressions commonly observed. According to Ekman (1992) an interpretation of facial expressions must rely on the postulated configurations and not on single facial actions.

There is considerable evidence (reviewed in Ekman, Friesen & Ellsworth 1982) indicating distinct prototypical facial signals that can be reliably recognized across a variety of cultures as corresponding to the emotions of happiness, sadness, surprise, disgust, anger, and fear. Figure 1 shows examples of some of these expressions.

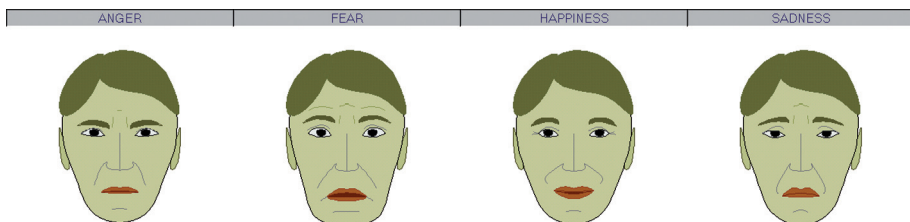


Figure 1. Prototypical expressions as postulated by discrete emotion theorists for the emotions of anger, fear, happiness, and sadness. The expressions shown have been synthesized with FACE (Wehrle 1995b/1999)

These patterns have been found in studies using photographs of posed facial expressions. However, these findings have not enabled researchers to interpret facial expressions as unambiguous indicators of emotions in spontaneous interactions. The task of analyzing the ongoing facial behavior in dynamically changing emotional episodes is obviously more complex than linking a static emotional expression to a verbal label.

Another problem not solved by the studies on the universal recognition of basic emotions concerns the dynamics of facial expressions. Generally, our theoretical knowledge of the temporal unfolding of facial expressions is quite limited. For example, one of the consistent findings in emotion recognition studies using static stimuli is that fear expressions are often confused with surprise. One explanation for this might be that the main difference between these two emotions resides in the respective temporal structure of the innervations in the facial musculature, which is not observable in still photographs. Another example of the relevance of temporal aspects is the fact that observers are very sensitive to false timing of facial expressions (e.g., abrupt endings or beginnings of a smile) when evaluating the truthfulness or deceitfulness of an emotional display. Although the importance of “correct” timing is widely accepted at a theoretical or phenomenological level, only a small number of studies have investigated this aspect systematically. The quantitative aspects of spontaneous facial expressions deserve further investigation, especially with regard to the duration of onset, apex, and offset.

The limitations of a basic emotion approach also have consequences for “affective” human-computer interactions and the above-mentioned goals of expression synthesis and expression decoding. Those prototypical full-face expressions might serve as icons representing basic emotions in simple emotional agents (synthesis) or as stereotyped signals of the user to indicate “anger” or “joy” (automatic expression recognition). Such signals could be used instead of more complicated and time-consuming messages via the keyboard like “I do not understand” or “this is fine with me”. However, a basic emotion approach does not allow us to make inferences about the emotional state of a user when interacting with a computer.

4.2 Componential appraisal theory and facial expression

Appraisal theorists following a *componential approach* share two main assumptions: (a) that emotions are elicited by a cognitive evaluation (*appraisal*) of antecedent situations and events, and (b) that the patterning of the reactions in the different response components (physiology, expression, action tendencies, and feeling) is determined by the outcome of this evaluation process. For these appraisal theorists the complexity and variability of different emotional feelings can

be explained without resorting to a notion of basic emotions. They argue that there are a large number of highly differentiated emotional states, of which the current emotion labels capture only clusters or central tendencies of regularly recurring ones, referred to as *modal* emotions. In line with this reasoning, facial expressions are analyzed as indicators of appraisal processes in addition to or as an alternative to verbal report measures. Facial expressions are not seen as the “readout” of motor programs but as indicators of mental states and evaluation processes. In contrast to discrete emotion theorists, they claim that single elements of facial patterns do have a meaning and that this meaning can be explained as manifestations of specific appraisal outcomes.

Several appraisal theorists have made concrete suggestions concerning possible links between specific appraisal dimensions and specific facial actions (Frijda & Tcherkassof 1997; Kaiser & Scherer 1998; Kaiser & Wehrle 2001; Smith & Scott 1997; Wehrle, Kaiser, Schmidt & Scherer 2000). Using the most recent version of FACS, Wehrle, Kaiser, Schmidt, and Scherer (2000) have extended and refined Scherer’s original predictions linking facial actions to the postulated appraisal checks (Scherer 1992). Scherer posits relatively few basic criteria and assumes sequential processing of these criteria in the appraisal process. The major “stimulus evaluation checks” (SECs) can be categorized into five major classes: (1) the novelty or familiarity of an event, (2) the intrinsic pleasantness of objects or events, (3) the significance of the event for the individual’s needs or goals, (4) the individual’s ability to influence or cope with the consequences of the event, including the evaluation of who caused the event (agency), and (5) the compatibility of the event with social or personal standards, norms, or values. As an example, Figure 2 shows the postulated facial expressions (Action Units) representing the appraisal profile for hot anger.

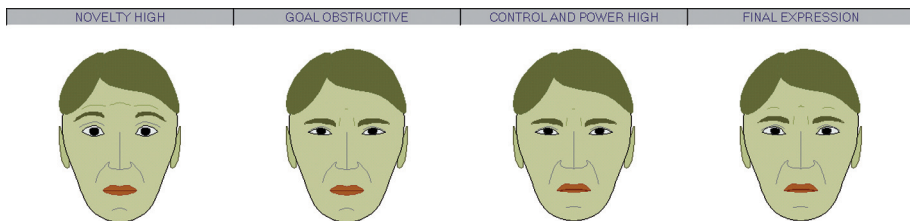


Figure 2. Predictions for the appraisal patterns and the related facial actions for hot anger as published in Wehrle *et al.* (2000). From left to right, the pictures illustrate the sequential cumulation of appraisal-specific facial Action Unit combinations resulting in a final pattern. Action Unit numbers and names: 1 (inner brow raiser), 2 (outer brow raiser), 4 (brow lowerer), 5 (upper lid raiser), 7 (lid tightener), 10 (upper lip raiser), 17 (chin raiser), 24 (lip press)

5. A computerized approach to the analysis and synthesis of facial expression

In spontaneous interactions, facial behavior accompanies emotional episodes as they unfold, and changes in facial configurations can occur very rapidly. In the preceding sections, we have illustrated some of the methodological and theoretical problems we are confronted with if we want to study the process of emotional interactions and its reflection in facial activity in interactive settings. To tackle some of these problems we are using a theory-based experimental paradigm including computerized data collection and data analysis instruments on the one hand, and computer simulation and theory-based synthetic stimuli on the other hand.

5.1 Experimental setup, data collection and data analysis

To study the dynamics and the interactive nature of emotional episodes, we developed the *Geneva Appraisal Manipulation Environment* (GAME; Wehrle 1996a), a tool for generating experimental computer games that translate psychological postulates into specific micro-world scenarios (for details about theoretical and technical embedding of GAME see Kaiser & Wehrle 1996; Kaiser & Wehrle 2001; Kaiser, Wehrle & Schmidt 1998). GAME allows automatic data registration of the dynamic game progress and the subjects' actions and automatic questionnaires. For example, participants' evaluations of specific situations are assessed by means of pop-up screens (which appeared after the completion of each game level) corresponding to 18 questions referring to Scherer's appraisal dimensions or SECs (see also Scherer 1993).

While playing the experimental game, participants are videotaped and these tape recordings allow an automatic analysis of the participant's facial behavior with the *Facial Expression Analysis Tool* (FEAT; Kaiser & Wehrle 1992; Wehrle 1992/1996). These facial data can be automatically matched with the corresponding game data using the vertical time code as a reference for both kinds of data. FEAT is a connectionist expert system that uses fuzzy rules, acquired from a FACS expert, to automatically categorize facial expressions. This formal expertise has been transformed into a network structure by a compiler program. The resulting network is then able to do the classification task, using FACS as the coding language (for more details see Kaiser & Wehrle 1992; Kaiser & Wehrle 2001). With FEAT we can precisely analyze the dynamics of facial behavior, including intensities and asymmetries.

Due to the complexity of our multimodal, process-oriented approach, the analysis and visualization of the different kinds of data becomes a challenge in

itself. The *Interactive Data Elicitation and Analysis* tool (IDEA; Wehrle 1996b) provides the possibility of analyzing multimedia behavioral records of an experimental session and to add secondary codings to the behavioral records. Behavioral records include all data registered during an experimental session, like videotaped data and data automatically registered in a protocol. TRACE (Wehrle 1999) can automatically reconstruct and analyze a played game in terms of topological measures (e.g., tactical constellations), critical episodes, urgency, goal multiplicity, etc. This facilitates a situational analysis of the behavioral records.

The ensemble of these components constitutes a situated coding procedure as defined in Section 3. In this case, the FEAT coding procedure is an integral part of the experimental setting, GAME. The interpretation of the automatically coded facial behavior is based on the simultaneously registered and automatically analyzed contextual and behavioral records, using IDEA and/or TRACE. In this way, the computer game provides a relatively simple but complete context for the interpretation of the internal emotional and cognitive regulatory processes. This is an important point because what the participant feels and what a certain facial expression means is often very context-specific.

5.1.1 *Some results: Facial expression and appraisal in human-computer interactions*

Within the scope of this chapter we can only present examples of the current research to illustrate the views on facial expression presented here (more details are published in Kaiser & Wehrle 1996; Kaiser & Wehrle 2001; Kaiser, Wehrle & Edwards 1994; Kaiser, Wehrle & Schmidt 1998; Wehrle & Scherer 2001; Wehrle *et al.* 2000). The evaluation of situational appraisal profiles allows us to differentiate between different types of identically labeled emotions. For example, in the experimental computer game interactions, we can consistently distinguish at least three types of anger: (1) being angry as an reaction to an unfair event but without blaming anybody, (2) being angry and blaming somebody else for having caused the event on purpose, and (3) being angry and blaming the other as well as oneself.

Figure 3 shows an example of anger of type 1. Here, the participant is on the seventh level of the game, which has started much faster than all levels she had played before. After a little while, Amigo – an animated agent that has the role of supporting and helping the player – intervenes, reducing the game speed to a “manageable” degree. In this situation, 73% of the participants report relief or happiness. However, this participant reports “anger”. As it can be seen in the appraisal profile, she evaluates the situation as very sudden, very new, very unpleasant, and not at all expected. Additionally, she thinks that, although the situation was difficult to control, she had enough power to handle it, that she can easily adjust to its consequences, and that her behavior (referred to as “Self” in Figure 3c) was

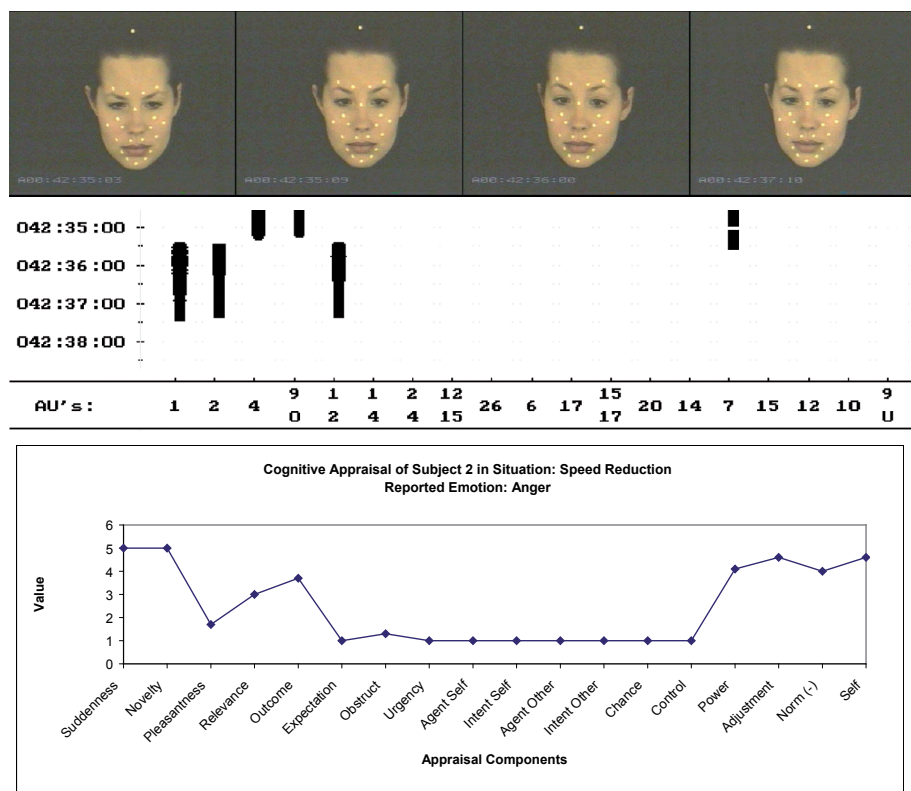


Figure 3a–c. A sequence of facial reactions that occurs in a situation in which AMIGO reduces speed. The top of the figure (a) shows still pictures of the subject’s face, the center (b) shows the results of the automatic FEAT coding, and the bottom (c) shows the participant’s evaluation of this situation in terms of Scherer’s appraisal dimensions (SECs) on Likert scales from 1 to 5. Figure 3b: The distribution of Action Units over a period of 4 seconds. The x-axis shows the repertoire of Action Units (AU) that are included in the knowledge base of the net. Similarly to the tracing of an electroencephalograph, the intensity of Action Units can be seen in the horizontal width of the bars. The onset and offset of an Action Unit as well as the duration of the apex can also be observed

adequate. She does not blame anybody or circumstances. However, she evaluates the situation (referred to as “Norm (–)” in Figure 3c) as being very “unfair”. As it can be seen in Figure 3, she reacts by raising her eyebrows (AU1 + AU2) only. Within a basic emotion approach, this Action Unit combination could not be linked to anger. When we look at the dynamics of the facial action, we see that a) the change from AU4 (she frowns while reading a message) and the innervation of AU1 and AU2 occurs within two frames (0.08 seconds). The duration of AU1 and AU2, however, is rather long – 2.04 seconds, whereas the mean duration of

AU1 and AU2 in these studies is 1.04 seconds. While the results show that “short” innervation of AU1 and AU2 is linked to the appraisal dimension of unexpectedness, in the case of the example shown in Figure 3, the holding of the expression can be interpreted as an indicator of appraising the situation as unfair. This interpretation is supported by the fact that naïve judges do recognize her nonverbal reaction as expressing anger. Although preliminary and speculative, the results show that we need to know more about the dynamics of facial expression.

5.2 Computer simulation and synthetic modeling

We complement our experimental studies with judgment studies that make use of synthetic faces and animated facial behavior. Such a synthetic approach has the advantage that the synthetic stimuli can be created on the basis of theoretical models. This allows one to compare and test even subtle differences between different theoretical models with respect to the dynamics and the variability of facial expressions. For this purpose we use the *Facial Action Composing Environment* (FACE), a tool for creating animated 3D facial expressions in real-time, including head and eye movements. The contours of the face are represented with splines, as are the prominent features of the face such as eyebrows and lips, but also wrinkles and furrows (see Figures 1 and 2). The repertoire of facial expressions for the animation is defined on the basis of FACS.

In addition to expression synthesis, we use computer simulation approaches for the development and testing of componential appraisal theories of emotion. The *Geneva Appraisal Theory Environment* (GATE; Wehrle 1995a) is a tool that allows the simulation of different appraisal theories as *black box models*. This tool enables us to refine or change theoretical propositions incrementally, and it provides immediate feedback on the outcomes. It can also quickly analyze existing data sets for an estimation of the new system’s global performance. GATE provides sophisticated facilities to systematically explore such empirical data sets. Furthermore, new empirical data can be obtained directly since the tool incorporates an automatic questionnaire that prompts a subject with a number of questions to determine the appraisal of a situation corresponding to a recalled emotional episode (for more details see Wehrle & Scherer 1995; Wehrle & Scherer 2001; Kaiser & Wehrle 2006).

Finally, we attempt to synthesize emotional behavior on the level of biologically plausible mechanisms. The *Autonomous Agent Modeling Environment* (AAME; Wehrle 1993) is a simulation environment for process modeling. It is a tool for designing autonomous agents for use in research and education. The intention was to have an adequate tool that helps to explore psychological and

Cognitive Science theories of situated agents, the dynamics of system- environment interactions, and the engineering aspects of autonomous agent design. In the case of the AAME the interesting aspects of such an agent are not so much the modeling of realistic sensors and effectors but the coupling mechanisms between them (for more details see Wehrle 1994a; Wehrle 1994b).

5.2.1 *Some results: Testing theoretical predictions with synthesized facial expressions*

Results from our studies on facial expression synthesis suggest that subjects perceive the synthetic images and animations generated by FACE in a similar fashion as photographs of real facial expressions (Wehrle *et al.* 2000). In addition, we used FACE to study the effect of static versus dynamic presentation of facial expressions. Here, we found that dynamic presentation increases overall recognition accuracy and reduces confusion. One explanation for this encouraging result might be that the facial repertoire was created on the basis of detailed descriptions of the changes in appearance produced by each Action Unit in the FACS manual. As mentioned above, we designed appearance changes not only for features like eyebrows but also to represent changes in the shape of the facial regions involved and the resulting wrinkles. As a consequence of this strategy, combinations of Action Units show the same appearance changes in the synthetic face as described in the FACS manual. These changes were not specified but *emerge* from adding the vector information of the respective single Action Units.

Further results of these judgment studies concern the differentiation of positive emotions for which discrete emotion theorists and appraisal theorists make different predictions. Proponents of discrete emotion theory argue that positive emotions such as amusement, contentment, excitement, pride in achievement, satisfaction, sensory pleasure, and relief, all share a particular type of smile – a Duchenne smile, which is defined as the combination of AU12 (lip raiser) and AU6 (orbicularis oculi) producing crow-feet wrinkles at the eye corners – and that no further facial signal will be found that differentiates among these positive emotions (Ekman 1994). Following an appraisal-based approach, Wehrle *et al.* (2000) postulate that these different positive emotions do produce different facial patterns, since the respective underlying appraisal profiles are not identical. The results show that judges differentiated well between sensory pleasure, happiness, elation and, to a lesser degree, pride. The predicted facial expressions for elation and sensory pleasure are shown in Figure 4.

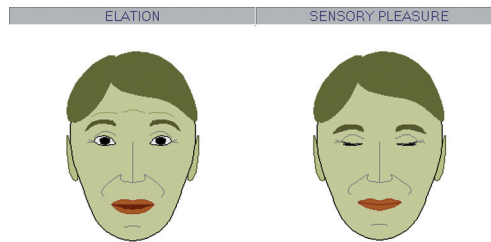


Figure 4. Predictions for the appraisal patterns and related facial actions of elation and sensory pleasure. Source: Wehrle *et al.* (2000)

6. Conclusion

This chapter has tried to show how an appraisal-based approach might help us to better understand how emotions are facially *expressed* and how facial expression is *perceived*. With respect to human-computer interaction, these two processes refer to the *synthesis* of “emotional agents” and the *analysis* of a user’s facial expression.

In our view, componential appraisal theory has some important theoretical, methodological, and ecological advantages over an emotion-based approach (see also Coulson, this volume). The appraisal-based approach does not only take into account the subjective evaluation of a situation that gives rise to an emotional experience but it directly links the outcome of this appraisal process to the other components of emotions. This allows us to analyze and synthesize emotional processes on a level of differentiation that goes beyond basic emotions. Since the process of appraisal and reappraisal is not only determined and changed by external situational cues but also by internal cues that reflect the person’s motives, personality, experience etc., these factors should also be considered explicitly.

From the beginning of the early eighties appraisal theories have given important inputs to emotion synthesis and affective user modeling (Elliot 1997). However, most of these applications only use appraisal theory to implement the “cognitive” component of an emotional interface (for reviews see Pfeifer 1988; Picard 1997). The outcome of the appraisal process is then mapped into an emotion category, which determines the synthesis of the respective facial expression pattern. This might be an unnecessary and complicated procedure that also reduces the available information and might bias analysis. Furthermore, linking facial expression and other components of emotion directly to appraisal dimensions can take advantage of the computational representations that are already defined in the respective computer applications.

Acknowledgements

This research was supported by grants from the Swiss National Science Foundation (FNRS 11-39551.93/049629.96 “Dynamic man-computer interactions as a paradigm in emotion research”) awarded to Susanne Kaiser and Thomas Wehrle.

References

- Ball, G., & Breese, J. (2000). Emotion and Personality in a Conversational Agent. In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied Conversational Agents* (pp. 189–219). Cambridge, MA: The MIT Press.
- Bänninger-Huber, E., & Widmer, C. (1996). A New Model of the Elicitation, Phenomenology, and Function of Emotions. In N. H. Frijda (Ed.), *Proceedings of the IXth Conference of the International Society for Research on Emotions* (pp. 251–255). Toronto: ISRE Publications.
- Cañamero, L. D. (Ed.) (2001). *Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium*. Technical Report FS-01-02. Menlo Park, CA: AAAI Press.
- Cañamero, L. (Ed.) (2005). *Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action. Papers from the AISB'05 Symposium*. SSAISB Press.
- Darwin, C. (1876/1965). *The expression of the emotions in man and animals*. Chicago: University of Chicago Press. (Original work published 1876 London: Murray)
- Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science*, 3, 34–38.
- Ekman, P. (1994). All emotions are basic. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 7–19). Oxford: Oxford University Press.
- Ekman, P., & Friesen, W. V. (1969a). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1, 49–98.
- Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). Research Foundations. In P. Ekman (Ed.), *Emotion in the Human Face* (pp. 1–143). New York: Cambridge University Press.
- Elliot, C. (1997). I Picked up Catapia and Other Stories: A Multi-modal Approach to Expressivity for ‘Emotionally Intelligent’ Agents. In W. L. Johnson (Ed.), *Proceedings of the First International Conference of Autonomous Agents* (pp. 451–457). New York: ACM Press.
- Essa, I., & Pentland, A. (1997). Coding, Analysis, Interpretation, and Recognition of Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 757–763.
- Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego: Academic Press.
- Frijda, N. H. (1986). *The emotions*. Cambridge and New York: Cambridge University Press.
- Frijda, N. H., & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 78–102). Cambridge: Cambridge University Press.
- Gigerenzer, G., & Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

- Hudlicka, E., & Cañamero, L. (Eds.) (2004). *Architectures for Modeling Emotion: Cross-Disciplinary Foundations. Papers from the 2004 AAAI Spring Symposium*. Technical Report SS-04-02. Menlo Park, CA: AAAI Press.
- Izard, C. E. (1991). *The psychology of emotions*. New York, NY: Plenum Press.
- Kaiser, S. (2002). Facial expressions as indicators of “functional” and “dysfunctional” emotional processes. In M. Katsikitis (Ed.), *The Human Face: Measurement and Meaning* (pp. 235–254). Dordrecht: Kluwer.
- Kaiser, S., & Scherer, K. R. (1998). Models of ‘normal’ emotions applied to facial and vocal expressions in clinical disorders. In W. F. Flack, Jr. & J. D. Laird (Eds.), *Emotions in Psychopathology* (pp. 81–98). New York: Oxford University Press.
- Kaiser, S., & Wehrle, T. (1992). Automated coding of facial behavior in human-computer interactions with FACS. *Journal of Nonverbal Behavior*, 16, 67–83.
- Kaiser, S., & Wehrle, T. (1996). Situated emotional problem solving in interactive computer games. In N. H. Frijda (Ed.), *Proceedings of the IXth Conference of the International Society for Research on Emotions* (pp. 276–280). Toronto: ISRE Publications.
- Kaiser, S., & Wehrle, T. (2000). Ausdruckspsychologische Methoden. In J. H. Otto, H. A. Euler, & H. Mandl (Hrsg.), *Handbuch Emotionspsychologie* (pp. 419–428). Weinheim: Beltz, Psychologie Verlags Union
- Kaiser, S., & Wehrle, T. (2001). Facial expressions as indicators of appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotions: Theory, methods, research* (pp. 285–300). New York: Oxford University Press.
- Kaiser, S., & Wehrle, T. (2006). Some studies for modeling the expression of emotion. In R. Trappl (Ed.), *Proc. Cybernetics and Systems 2006* (pp. 619–624). Vienna: Austrian Society for Cybernetic Studies.
- Kaiser, S., Wehrle, T., & Edwards, P. (1994). Multi-Modal Emotion Measurement in an Interactive Computer Game: A Pilot-Study. In N. H. Frijda (Ed.), *Proceedings of the VIIIth Conference of the International Society for Research on Emotions* (pp. 275–279). Storrs: ISRE Publications.
- Kaiser, S., Wehrle, T., & Schmidt, S. (1998). Emotional episodes, facial expression, and reported feelings in human-computer interactions. In A. H. Fischer (Ed.), *Proceedings of the Xth Conference of the International Society for Research on Emotions* (pp. 82–86). Würzburg: ISRE Publications.
- Lazarus, R. S. (1966). *Psychological Stress and the Coping Process*. New York: McGraw Hill.
- Leventhal, H., & Scherer, K. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion*, 1, 3–28.
- Lien, J. J., Kanade, T. K., Zlochow, A. Z., Cohn, J. F., & Li, C. C. (1998). A Multi-Method Approach for Discriminating Between Similar Facial Expressions, Including Expression Intensity Estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Santa Barbara, CA.
- Paiva, A. (Ed.) (2000). *Affective Interactions: Towards a new generation of computer interfaces*. Heidelberg: Springer.
- Paiva, A., Prada, R., & Picard, R. (Eds.) (2007). *Affective Computing and Intelligent Interaction, Second International Conference, ACII 2007* [LNCS 4738]. Berlin/Heidelberg: Springer.
- Pelachaud, C. (2005). Multimodal Expressive Embodied Conversational Agents. In H. Zhang, T.-S. Chua, R. Steinmetz, M. Kankanalli, & L. Wilcox (Eds.), *Proceedings of the 13th annual ACM international conference on Multimedia* (pp. 683–689). New York, NY: ACM Press.

- Pelachaud, C., & Cañamero, L. (Eds.) (2006). Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids. Special issue of the *International Journal of Humanoid Robotics*, 3(3).
- Pfeifer, R. (1988). Artificial Intelligence Models of Emotion. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive Perspectives on Emotion and Motivation* (pp. 287–320). Dordrecht: Kluwer Academic Publishers.
- Picard, R. W. (1997). *Affective Computing*. Cambridge: The MIT Press.
- Roseman, I., & Kaiser, S. (2001). Applications of appraisal theory to understanding and treating emotional pathology. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal theories of emotions: Theories, methods, research* (pp. 249–270). New York: Oxford University Press.
- Russell, J. A., & Fernández-Dols, J. M. (Eds.) (1997). *The psychology of facial expression*. Cambridge: Cambridge University Press.
- Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–318). Hillsdale, NJ: Lawrence Erlbaum.
- Scherer, K. R. (1992). What does facial expression express? In K. Strongman (Eds.), *International Review of Studies on Emotion*. Vol. 2 (pp. 139–165). Chichester: Wiley.
- Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition and Emotion*, 7, 325–355.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813–838.
- Smith, C. A., & Scott, H. S. (1997). A componential approach to the meaning of facial expression. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 229–254). Cambridge: Cambridge University Press.
- Takeuchi, A., & Naito, T. (1995). Situated Facial Displays: Towards Social Interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 450–454). New York: ACM Press.
- Tao, J. T. Tan, & Picard, R. (Eds.) (2005). *Affective Computing and Intelligent Interaction, First International Conference, AII 2005* [LNCS 3784]. Berlin/Heidelberg: Springer.
- Wehrle, T. (1992/1996). *The Facial Expression Analysis Tool (FEAT)* [Unpublished computer software]. University of Geneva, Switzerland.
- Wehrle, T. (1993). *The Autonomous Agent Modeling Environment (AAME)* [Unpublished computer software]. University of Geneva, Switzerland.
- Wehrle, T. (1995/1999). *The Facial Action Composing Environment (FACE)* [Unpublished computer software]. University of Geneva, Switzerland.
- Wehrle, T. (1994a). *Eine Methode zur Psychologischen Modellierung und Simulation von Autonomem Agenten* (A method for the psychological modeling and simulation of autonomous agents). Unpublished doctoral dissertation, University of Zürich.
- Wehrle, T. (1994b). New Fungus Eater Experiments. In P. Gaussier & J.-D. Nicoud (Eds.), *From Perception to Action* (pp. 400–403). Los Alamitos: IEEE Computer Society Press.
- Wehrle, T. (1995a). *The Geneva Appraisal Theory Environment (GATE)*. [Unpublished computer software]. University of Geneva, Switzerland.
- Wehrle, T. (1995/1999). *The Facial Action Composing Environment (FACE)* [Unpublished computer software]. University of Geneva, Switzerland.
- Wehrle, T. (1996a). *The Geneva Appraisal Manipulation Environment (GAME)* [Unpublished computer software]. University of Geneva, Switzerland.

- Wehrle, T. (1996b). *The Interactive Data Elicitation and Analysis (IDEA)* [Unpublished computer software]. University of Geneva, Switzerland.
- Wehrle, T. (1999). *Topological Reconstruction and Computational Evaluation of Situations (TRACE)*. [Unpublished computer software]. University of Geneva, Switzerland.
- Wehrle, T. (2001). The grounding problem of modeling emotions in adaptive artifacts. In P. Peta & L. D. Cañamero (Eds.), *Grounding emotions in adaptive systems: Volume I* [Special issue]. *Cybernetics and Systems: An International Journal*, 32(5), 561–580.
- Wehrle, T., & Scherer, K. R. (1995). Potential Pitfalls in Computational Modelling of Appraisal Processes: A Reply to Chwelos and Oatley. *Cognition and Emotion*, 9, 599–616.
- Wehrle, T. & Scherer, K. R. (2001). Towards computational modeling of appraisal theories. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotions: Theory, methods, research* (pp. 350–365). New York: Oxford University Press.
- Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression via synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78, 105–119.

Expressing emotion through body movement

A component process approach

Mark Coulson

1. Introduction

The human body represents a potentially important medium for the expression of emotion. Bodies are large, infinitely poseable and easily perceived, and as such might be considered one of the primary means by which emotions are communicated. However, body movement and posture has not received the same degree of attention that facial and vocal expression have attracted, and remains the 'poor relation' in the family of emotional expression and communication. Our understanding of bodily expression of emotion remains limited (see e.g. Bianchi-Berthouze & Kleinsmith 2003; Camurri *et al.* 2005; Coulson 2004; Kleinsmith *et al.* 2006) despite the obvious uses of animated characters for instruction, communication and entertainment.

There are several reasons for this lack of knowledge and research. First, bodies are complex objects which exhibit a large number of degrees of freedom (defined as joint rotations about one or more axes) and an enormous variety of postures and movements. The systematic investigation of the resulting high-dimensional space is a considerable challenge, and although several attempts at dimension reduction have been made (see for example Montepare *et al.* 1999, dimensions of form, tempo, force and direction), none of these has been related to actual joint rotations, and their precise implications are unclear. Hence there exists a gap between perceptual gestalts and actual bodily configurations or patterns of movement which hampers attempts at simulation.

Second, the primary functions of bodies are locomotion and manipulation rather than expression. Despite being capable of expressing emotion, bodies have not evolved to do so. Whereas movements of the facial musculature may be directly interpreted as expressive, inferring emotion from body movement is likely to be a more indirect process. Movements are functional actions and responses to stimuli and situations which may or may not have emotional consequences. The affective content of a sequence of body movements is therefore perceived as

epiphenomenal rather than as a direct expression of the emotion. Also, as movements which may be considered expressive are embedded within ambulatory or manipulative ones, the signal to noise ratio of bodily expression of emotion is poor relative to facial and vocal channels. As pointed out by Kaiser & Wehrle (this volume), the signal to noise ratio for facial expressions is itself not all that high, but for bodies this is further confounded.

Third, there is a great deal of individual variation in the way in which people express and perceive emotion from body movement (Wallbott 1998). While there is considerable evidence that the vast majority of people produce very similar facial expressions when experiencing emotion, the same has not generally been found for body posture. If it were the case that bodies did express emotion directly but this was done in different ways by different people, then the shared meanings which are essential to any system of communication would not be present, and people would not reliably infer emotion in each other on the basis of body movement. If, in addition, people perceive emotion from posture and movement idiosyncratically, the whole endeavour becomes even more problematic.

Fourth, 'body language' is generally seen as more culturally distinct than universal, and as such the meaning of postures and movement is held to vary across cultures to a degree which makes universal recognition unlikely. In the face of tremendous success in the field of universal recognition of facial expression, body movement appears to be a less attractive domain for simulation. Again, the extent to which this is true has serious implications for the effective scope of simulations.

Despite these challenges, the simulation of expressive movements has application to a wide variety of fields and represents a worthy pursuit. If we can start to specify how different types of movement lead to inferences about different emotions, we can generate more convincing and acceptable animations for use in education and entertainment, we can offer useful advice to those working in professions which value any insight into a client's mental state, and we can educate and train people to, perhaps quite literally, 'put their best foot forward' in personal and professional interactions.

Before describing how emotional animations are created, however, one final problem needs to be addressed, that is the nature of emotion. The next section outlines theories of emotion, and identifies how adopting a particular perspective may overcome some of the difficulties arising from individual and cultural variation in expression. Following the theoretical discussion I outline the nature of the implementation and conclude by discussing challenges and outstanding issues.

2. A theoretical basis for emotional expression

A theory of emotion is a necessary pre-requisite for any attempt to simulate emotional expression. At the very least we need to decide what emotion is, what causes it, and how it relates to expression. As will be seen, the theoretical position adopted here permits emotional expressions to be simulated without the necessity for explicitly identifying a person's emotion in terms of lexical tokens.

Three general classes of emotion theory are briefly outlined, following which a more detailed account of the approach adopted here is provided. Although the descriptions omit a great deal of detail, they capture the essence of the competing theoretical approaches. In each, it is assumed that the nature, experience and expression of emotion are determined by the same set of theoretical constructs.

2.1 The classical view of emotions

Emotions are qualitatively distinct states which are adequately described by linguistic tokens – fear, anger, sadness etc. Phenomenological experience, physiological response, preparedness for action and expressive behaviours are closely associated as patterns of response to stimuli or events. Emotions are associated with specific expressions, although these may be modified or inhibited according to personal, situational, and cultural standards. This position is most closely associated with researchers such as Paul Ekman (e.g. Ekman 1993) who in a series of studies over the past 30 years has convincingly demonstrated that facial expressions for at least six 'basic' emotions (anger, disgust, fear, happiness, sadness and surprise) are recognised at similarly high levels in a wide variety of cultures.

There are a number of areas where the classical approach to emotion has been criticised, not all of which are relevant here. One point which is often overlooked is that the lexical items which are used to identify emotions do not make any predictions about how a person experiencing a particular emotion will respond. Knowing that someone is afraid tells us little about how they are going to respond. Rather, we need to know what has caused the fear, whether the threat is immediate or potential, whether it is to the physical or the social self, and so on. Indeed, as will be outlined later, once we have answers to these more basic but detailed questions, knowing that the person is experiencing fear becomes redundant.

2.2 The dimensional view of emotions

Emotions result from the combination of two or more orthogonal dimensions. The first two are generally identified with valence and arousal (Reisenzein 1994),

or rotated through 45 degrees and labelled positive and negative affect (Watson, Wiese, Vaidya, & Tellegen 1999). Both variants result in a circumplex structure, with emotion tokens occupying locations on the circumference of a circle bisected by the two dimensions. Emotion tokens are valid descriptions of emotion, but do not represent qualitatively different categories or states, being instead the result of varying combinations of an underlying set of dimensions.

The dimensional view suffers from some of the same limitations as the classical view in that it offers no simple way of prescribing appropriate responses to emotional events. It has the advantage of being somewhat more fine-grained in that emotions are decomposed into a number of simpler dimensions, but specifying the effect different levels of these may have on movement is difficult. There is also considerable debate over the ontological status of valence and arousal. While the former has been strongly linked with approach and avoidance behaviour (see, for instance, Carver & Scheier 1990), the arousal dimension remains controversial. Translating dimensional values into body movements is not possible with anything approaching a realistic level of detail, and a more fine-grained theory is required.

2.3 The component process view of emotions

Emotions are the end result of a linear and cyclic series of cognitive appraisals. The appraisal system continually evaluates the environment in terms of its significance for the organism's ongoing activity. Appraisal outcomes have immediate and direct effects on all aspects of the organism, and the temporal unfolding of these effects as they pertain to the face, the voice and the body constitutes emotional expression. Appraisal outcomes determine expression and emotion separately – expressions are the moment by moment unfolding of behavioural responses to appraisals while emotions are a phenomenological experience of the entire series of appraisals.

The heart of component process views is not emotion and experience, but cognition and evaluation. Complex appraisal theories view emotion as the affective significance of a series of evaluations. When we appraise a stimulus as being inherently unpleasant with immediate and negative consequences for our self, and upon which we are powerless to act, we feel fear or despair. The emotion is the end result of the series of appraisals, which plays no causal role in the genesis of behaviours mandated by those appraisals. Although certain aspects of emotion are captured by linguistic tokens, these are only rough approximations.

By breaking down the process of emotion generation to a more fundamental cognitive level, component process theories can be used to identify the physical and behavioural implications of emotional events as they occur. What follows

is a description of an implemented component process approach to the bodily expression of emotion.

3. Applying a component process approach

The most carefully thought through version of the component process approach is that of Klaus Scherer (1987, 2001). Under Scherer's model there are four main objectives of the appraisal system. These objectives concern the *Relevance* of the stimulus to the appraising individual, its *Implications*, the *Coping* abilities of the individual, and the effect of the appraised stimulus on a variety of *Normative Standards*. These objectives provide a structure within which individual appraisals are made. Each objective is subdivided into a number of specific appraisals, referred to as Stimulus Evaluation Checks (SECs) see Table 1. Evaluation of a stimulus proceeds by processing each SEC in turn, cyclically, so that once the sequence finishes it starts again from the beginning. Experience, behaviour and expression arise independently from this sequence (see Figure 1).

As an example, consider the appraisal of a sudden loud noise in the middle of the night. According to Scherer's theory, the three SECs covering the Relevance appraisal objective process the stimulus as novel, intrinsically unpleasant, and

Table 1. Thirteen stimulus evaluation checks organised under four appraisal objectives

Appraisal objective	SEC	Question SEC answers
Relevance	Novelty	<i>Has the environment changed in a significant way?</i>
	Intrinsic pleasantness	<i>Is the stimulus in and of itself good or bad?</i>
	Goal relevance	<i>Does it affect my ongoing activity?</i>
Implications	Causal attribution	<i>Who or what has caused it to happen?</i>
	Outcome probability	<i>How likely are the various outcomes of this change?</i>
Implications (cont.)	Expectation discrepancy	<i>Does what has happened fit with my expectations?</i>
	Goal/need conduciveness	<i>Does this help or hinder me?</i>
Coping	Urgency	<i>Is this something which has to be dealt with now?</i>
	Control	<i>Can it be controlled (in principle)?</i>
	Power	<i>Am I able to control it?</i>
	Adjustment	<i>Can I adjust to the change it entails?</i>
Normative significance	Internal standards	<i>Does this fall short of, within, or exceeds my own internal standards?</i>
	External standards	<i>Does this fall short of, within, or exceeds my own external standards?</i>

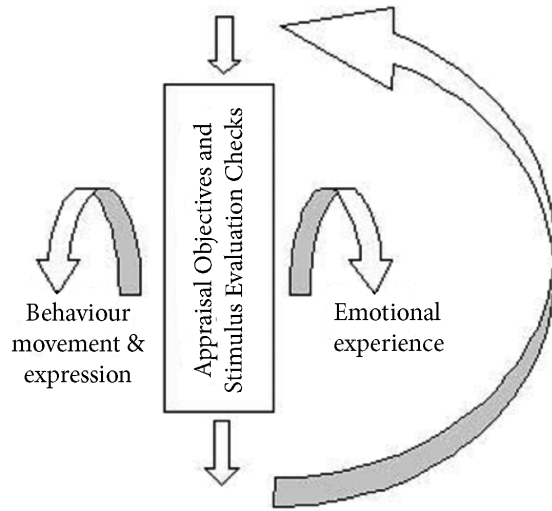


Figure 1. The linear and cyclic process of stimulus evaluation

relevant to ongoing activity (Novelty check, Intrinsic Pleasantness check, Goal Relevance check). In other words, the stimulus is something which is new, which is not nice, and which is important to the individual's ongoing concerns, namely having a good night's sleep.

The Implications objective appraises the stimulus as being caused by an external agent or event, with an as yet unknown likelihood that action will be required (Causal Attribution check, Outcome Probability check). Furthermore, the stimulus is appraised as discrepant from expectations, hindering rather than helping, and possibly urgent (Expectation discrepancy check, Goal/need conduciveness check, Urgency check). Depending on the nature of the noise and on the individual hearing the noise, Coping Potential SECs might classify the stimulus as being potentially controllable, and the individual may not perceive him/herself as having the power to enforce this control (Control check, Power check). He/she will feel more or less capable of adjusting to the stimulus without further action (Adjustment check). Finally, the Normative Standards checks determine whether the stimulus violates expected codes of personal (internal) or external behaviour. In the scenario outlined above, the cause of the event is external to the individual appraising its significance, and these checks will not be required. As each check returns a result, there is the potential for a direct effect on the individual's behaviour and expression. Should the result of an SEC specify that a particular course of action is sensible in the current appraisal context, then that action will be instigated immediately. The primary effect of SECs is therefore to direct behaviour.

A secondary and cumulative (as opposed to componential) effect of the sequence of SECs is an emotion experience which one might describe as fear. Note, however, that the expression of this emotion has developed independently from the emotion. The emotion does not create the expression but is instead the experience of the appraisal results which give rise to the expression.

Several important points arise from this example. First, emotions arise as the cumulative and interactive effect of all SECs, and are the result of complex patterns of appraisal. They are therefore not categorical states. They are also not simply the result of an underlying dimensional structure. Although a number of appraisals are modelled as scalar values, others return categorical results. The conceptualisation of emotion remains distinct from both categorical and the dimensional theories.

Second, there is no point at which the cycle of processing is complete as stimuli are constantly re-evaluated and initial outcomes may be revised as more information becomes available. The dynamic essence of emotion is therefore captured well.

Third, not all SECs necessarily contribute to all emotions. In the example given above, violations of Normative Standards are not relevant to fear responses, and these SECs remain open or even unprocessed. Likewise, something which is novel, inherently pleasant, and conducive to ongoing activity is likely to make us happy regardless of whether it is consistent or discrepant with our expectations, and whether we have the power to do anything about it.

Finally, although a SEC may not be relevant to the core definition of an emotion such as fear, it may still have an effect on expression and experience. For instance, the Adjustment SEC is open for 'fear' in Figure 2, but a high value for this would reflect adaptive fear (scared, but able to do something about it) whereas a low value would refer to something closer to blind terror (scared and powerless). The lexical token 'fear' captures the essence of the experience, but the full account lies only in the pattern of SEC outcomes. Patterns of appraisal outcomes and not lexical tokens give emotions their meanings.

4. Implementing a component process approach

The result of each SEC may have an immediate effect on all aspects of the organism as it responds to the demands of the appraisal. In terms of body movement, SEC outcomes specify bodily configurations which represent functional responses to the current appraisal of the stimulus, or the *appraisal context*. The bodily configurations specified by SEC outcomes are referred to as *target postures* due to their specifying a postural response which should be adopted following a

specific appraisal outcome. A positive novelty check, for instance, indicates there is something new and potentially important in the environment which should be attended to, and the target posture should satisfy this need. In this case, a positive novelty check results in an orienting response – weight transfer is slightly backwards, head and upper torso turn to face the stimulus. This response is a direct effect of the appraisal.

As the appraisal sequence unfolds, the target posture changes as more SEC results are returned, and the body will move to adopt the relevant target posture. It is assumed that speed of appraisal is generally faster than the body's rate of movement, so the target posture will not generally be attained before being superseded by subsequent SEC outcomes. The result of this is a continuously moving body trying to keep up with the demands of the incoming stream of appraisal outcomes.

From an implementation point of view the task is to identify those SEC outcomes which lead to functionally significant postures, model these postures using an appropriate three-dimensional representation of a human body, and code the expressive response as a dynamic series of transitions between target postures. Each of these issues is addressed below.

4.1 SEC outcomes and target postures

A target posture can be defined as a configurational solution to the cognitive-affective demands made on an individual by the appraisal context. These demands serve cognitive and informational functions (for instance leaning backwards widens the perceptual field and leaning forwards narrows it) as well as simple affective ones (positive appraisals result in approach movements, negative ones in withdrawal).

For early SECs, the process of determining a target posture is relatively straightforward. As mentioned above, an appropriate response to a positive Novelty check is orienting. Any other outcome of the Novelty SEC results in no change to the current posture. Something is either novel, in which case it should be attended to, or it is not, in which case current activity should not be interrupted.

For SECs occurring further downstream in the process, the situation becomes somewhat more complex. The response to a positive Relevance check will clearly depend on whether the stimulus is inherently pleasant or unpleasant. Indeed, the valence of the stimulus permeates almost all subsequent SECs, and the general principle is that the target postures associated with any specific appraisal are determined by the cumulative appraisal context and not just by the current SEC.

The interdependencies become yet more complex for later SECs. Responses to the Power check, which determines whether the individual is capable of do-

Table 2. Three example target postures

SEC	Target posture	Description
<i>Early SEC</i> Positive Novelty check		Body turns to face stimulus, chest leans back to broaden perceptual field. Arms raised in preparation for any required action.
<i>Middle SEC</i> Negative Control, Unpleasant, caused by another Agent		Head bowed to avert eyes, arms allowed to fall to the side, weight transfer neutral, body sags forward.
<i>Late SEC</i> Ought self (Standards check) exceeded		Head is up, arms relaxed to the sides, body erect.

ing anything about, or to, the stimulus, depends on the outcomes of three other checks. How we respond to the realisation that we can do something about a situation depends crucially on what the situation is (Intrinsic pleasantness check), what or who caused it (Causal Agent check), and whether it is in principle something which can be changed (Control check). The design of postures which represent such complex outcomes is not an exact science, and continual refinement is a necessary feature of this work.

Currently, a total of 24 distinct target postures are specified, each of which relates to one or more SEC outcomes. As theoretical analysis of the relationship between SECs and postures develops, this number is likely to increase. By way of example, three target postures from an early, an intermediary, and a late SEC are presented and briefly described in Table 2.

4.2 A model of the human body

Human bodies are complex physical objects exhibiting a large number of degrees of freedom and infinite variety in the postures they can adopt. Some simplifying assumptions therefore need to be made in order to produce a model which is tractable yet which offers sufficient complexity to deal with the full range of expressive movements.

The model used to generate the animations consists of 17 segments whose relative positions are described by 54 variables representing rotations about the major joints. The BVH format is used to represent these variables (see www.biovision.com/bvh.html for details) as this is simple to manipulate and can be read into

Curious Labs' POSER, the package used to generate animations. In this format, the hips form the top level of a hierarchy with all other segments attached directly or indirectly. Once the three dimensional location of the hips has been defined, the positions of the other segments are a function of this plus the specific joint rotations affecting each segment. For example, the position of a hand is defined by the location of the hips plus rotations about the waist (twist and bend), chest (bend and lean), shoulder (twist, adduct/abduct, forwards/backwards), and elbow (twist and bend).

4.3 Details of the implementation

All SECs are input to the program by the user or read in as previously defined patterns. The pattern of outcomes may be designed to represent a specific emotion, or may be any arbitrary set. The time at which each SEC triggers may also be specified, with the restriction that SECs must occur in linear sequence. Scherer argues that each SEC must reach 'preliminary closure' before the next SEC can begin processing the stimulus, so some flexibility in terms of when each SEC triggers is required.

Certain SECs have scalar values (e.g. Expectation Discrepancy, Goal Relevance, and Urgency) whereas others are qualitative categories (e.g. Causal Attribution, which is either the self, another agent, or an event or object, see Ortony, Clore & Collins 1988). The magnitude of a scalar value may influence the nature of the posture and may also affect the speed with which the figure moves. For instance, a highly unpleasant and discrepant stimulus is likely to result in a more rapid response than one which is only mildly unpleasant and slightly discrepant. Although both will tend towards the same target posture as outcome, the former situation will be associated with more rapid movement. SEC outcomes can therefore qualitatively specify a target posture towards which the body will move and/or quantitatively affect the speed of movement.

Once the pattern and timing of SECs has been defined and parameters describing the target postures for all SEC outcomes read in, the program generates 100 frames of animation representing smooth interpolations towards the sequence of target postures specified by the unfolding appraisal context.

The output is then imported into POSER, and animations produced from this. A static representation of the target postures involved in generating the 'fear' expression described above is presented in Figure 2, and Figure 3 shows a sample of frames taken from the animation produced by this sequence. Taken together these figures provide an indication of how the target postures 'drive' the animation, and how the animation moves towards an ever-changing end point.

Relevance SECs (outcome)	 Novelty (<i>positive</i>)	 Intrinsic pleasantness (<i>negative</i>)	no new posture specified Goal relevance (<i>relevant</i>)
↓			
Implications SECs (outcome)	no new posture specified Causal attribution (<i>open</i>)	no new posture specified Outcome probability (<i>high</i>)	 Expectation discrepancy (<i>high</i>)
↓			
SECs (outcome)	 Goal/need conduciveness (<i>low</i>)	effect of high urgency is to increase subsequent speed of movement Urgency (<i>high</i>)	
↓			
Coping potential SECs (outcome)	 Control (<i>mid to high</i>)	 Power (<i>low</i>)	no new posture specified Adjustment (<i>open</i>)
↓			
Normative standards SECs (outcome)	no new posture specified Actual self (<i>open</i>)	no new posture specified Ideal self (<i>open</i>)	no new posture specified Ought self (<i>open</i>)

Figure 2. A series of target postures for a sequence of appraisals leading to fear. Arrows indicate the progress of the appraisal. The target posture in each cell is that towards which the body will move following preliminary closure of the SEC

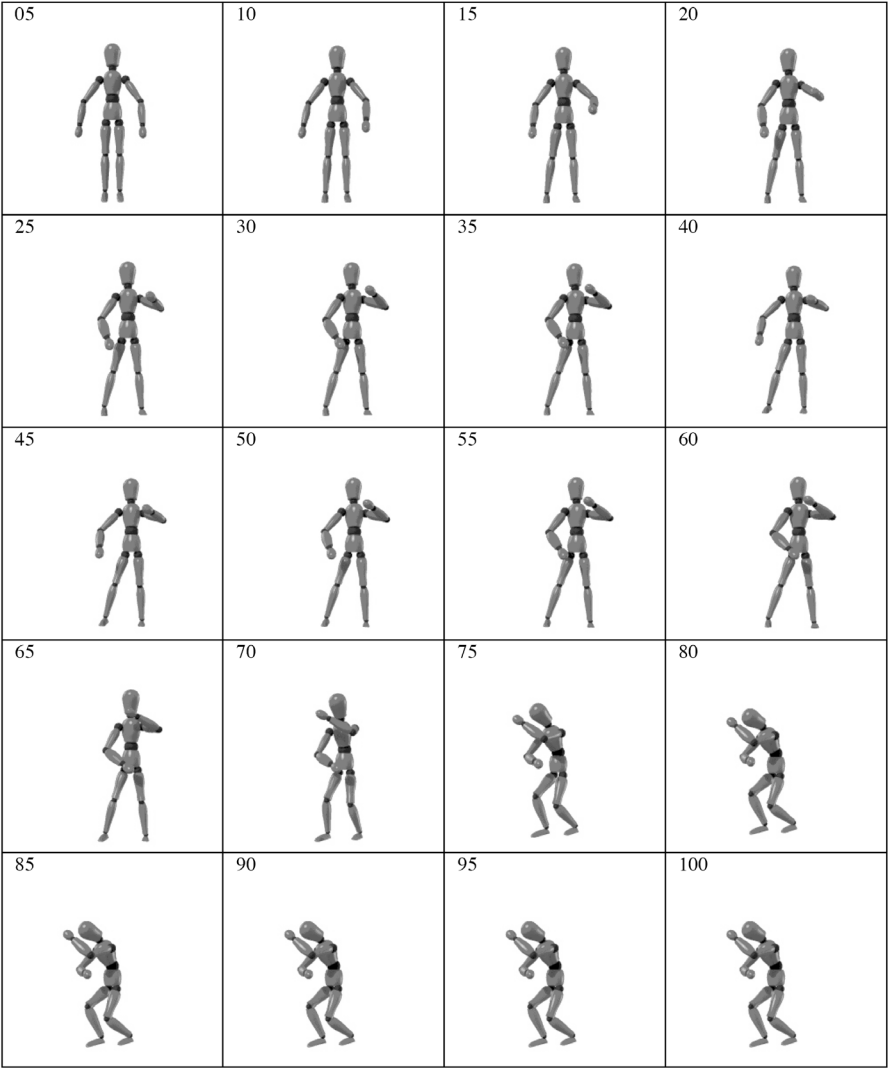


Figure 3. Representation of the animation produced from the series of SEC outcomes outlines in Figure 2 and in the text. Numbers refer to frame position in a 100-frame animation

5. Challenges

The current version of the implementation provides a starting point for the evaluation of both the component process approach and the effectiveness of the modelling. Three main challenges present, and each is examined below.

5.1 Specification of functionally significant postures resulting from SEC outcomes

The exact posture specified by each SEC outcome is determined in terms of its functional significance, and a given series of SEC outcomes will always specify the same posture. This is a limitation in the model's flexibility as a single specific posture is never the only functionally appropriate response to an appraisal. However, the ever-changing nature of the response to a stimulus, and the effect that the stimulus itself has on movement may militate against this being a serious restriction to the implementation. A more complete model will need to embed SEC-driven movements within those specified by both ongoing activity (walking, drinking a cup of coffee, etc.), and the emotion-eliciting stimulus itself (aggressive movements will vary according to the height of the person we are aggressing towards). The key to simulating realistic expressive movements may lie in loosely integrating multiple constraints rather than specifying one to an unnecessarily fine degree.

More difficult is the production of appropriate postures for SEC outcomes further down the chain of appraisal. To the extent that responses to later appraisals depend on earlier ones, the number of potential postures propagates. While this may reflect the complexity and flexibility of normal response, it requires careful analysis and the continual refinement of postures.

Finally, the current model only represents a single pass through the sequence of SECs. As appraisal is an ongoing cyclic process, a more complete model would need to incorporate this and allow for multiple passes through the sequence. This raises a secondary question about closure and how repeated passes through the appraisal sequence impact on the current posture. While it may be sufficient to assume that postural change will only arise as a consequence of a change in the appraisal context, this issue remains to be investigated.

5.2 Transitions between states: Realistic movement and inverse kinematics

Currently all transitions between successive postures are handled by smoothly interpolating joint rotations from current to target posture across a set number of

frames of animation. While this offers a simple means of creating animations and experimental stimuli, it is an oversimplification for two reasons.

First, speed of movement as defined by angular velocity about an axis of rotation does not remain constant across a typical movement. Onset tends to be rapid and offset more gentle, although this will vary with the urgency of the movement. Change in movement is also dependent on the momentum of the current movement, which is a function of its speed and the mass of the body segment being moved (it is easier to deflect a forearm than a thigh). A more realistic model will need to incorporate these and other facts if it is to generate realistic and meaningful animations.

Second, transfer of the mass centre is generally achieved by lifting the foot and stepping forwards or backwards. This action is relatively slow to initiate and may arise as a consequence of other movements which affect the body's balance as opposed to being an integral part of them as currently modelled. Sequencing movements and modelling the nature of weight transfer more realistically will clearly add to the naturalness of the animations.

5.3 Adequacy of theoretical assumptions

The component process approach is an ambitious and complex attempt to explain the cognitive precursors of emotion, and repeated testing of its predictions and assumptions can only strengthen the model. The approach has been shown to have utility in studies of facial (Wehrle, Kaiser, Schmidt, & Scherer 2000) and vocal (Banse & Scherer 1996) expression, and the framework has been refined and developed over the past 15 years. Its detail and complexity offers an excellent basis for the investigation of emotional expression through body movement.

In a complementary fashion, the process of simulation offers a means of testing the adequacy of theoretical assumptions. The current implementation models the Causal Attribution SEC in terms of Ortony *et al.*'s (1988) analysis of the cognitive structure of emotions, as this SEC is somewhat underspecified in the current version of the theory. Additionally, Normative Standards are modelled as arising from three models of the self: the actual, the ideal and the ought self (Higgins 1987). Rather than distinguishing between internal and external normative standards, as it is the case in the current version of Scherer's theory, the model views standards as equalling, exceeding, or falling short of the actual, ideal and ought selves. Higgins has argued that patterns of discrepancy between the different notions of self give rise to a variety of emotional responses, and although equating self-discrepancy with emotional labels does not map well onto the component process approach, her model of how patterns of discrepancy reflect affective pro-

cesses provides an excellent framework for modelling the relationship between appraisal and expression.

The interplay between theory and simulation is further strengthened by experimental work which examines the responses of human participants to animations. Some preliminary data indicate that recognition rates across a dozen distinct emotion tokens range from chance levels for tokens such as despair and disgust, to reasonable degrees of agreement (around 40–50% in a twelve-alternative forced-choice decision task) for anger, pride, fear and happiness. The triangulation of results from theoretical, simulation and experimental analyses provides a strong means by which all three can produce increasingly realistic accounts of how movement communicates emotion.

6. Conclusions

The strength of the component process approach lies in its definition of expression in terms of an individual's evaluation of a situation as opposed to a direct effect of an imprecisely defined lexical token. Expressions can therefore be simulated as a direct consequence of cognitive appraisals and, although this rests upon the theoretical view of emotion outlined above, the generation of animated affective characters does not need to simultaneously model those characters' actual emotions. If a stimulus or event can be described in terms of the SEC outcomes that it provokes, then an emotionally expressive response can be generated. Modelling these SEC outcomes may be a simpler task than modelling emotions. Indeed, removing emotion words from the analysis of expression may remove an artificial constraint on the degree to which expressions across all channels appear to be individually and culturally dependent. As Wierzbicka (1992) has persuasively argued, the use of emotion words to analyse emotion may be seriously misguided, and an approach based on utilizing fundamental components such as the SECs described by Scherer offers a way of avoiding this. Ongoing development of the implementation is focussing on the practical, biomechanical and theoretical aspects outlined above in addition to testing human participants' attributions of emotion to the model's output.

Acknowledgements

I should like to acknowledge the assistance and advice of Dr Rebecca Gould and Stephen Nunn in the preparation of this paper.

References

- Banse, R. & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636.
- Bianchi-Berthouze, N. & Kleinsmith, A. (2003). A categorical approach to affective gesture recognition, *Connection Science*, V(15-4), 259–269.
- Camurri, A., De Poli, G., Leman, M., & Volpe, G. (2005). Toward communicating expressiveness and affect in multimodal interactive systems for performing arts and cultural applications, *IEEE Multimedia*, 12(1), 43–53.
- Carver, C., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, 97 19–35.
- Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, 28(2), 117–139.
- Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, 48, 384–392.
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, 94, 319–340.
- Kleinsmith, A., De Silva, R., & Bianchi-Berthouze, N. (2006). Cross-Cultural Differences in Recognizing Affect from Body Posture. *Interacting with Computers*, 18(6), 1371–1389.
- de Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13, 247–268.
- Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior*, 23(2), 133–152.
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.
- Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, 67, 525–539.
- Scherer, K. R. (1987). Toward a dynamic theory of emotion: The component process model of affective states. Unpublished manuscript, University of Geneva.
- Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, Methods, Research*. New York & Oxford: Oxford University Press.
- Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28(6), 879–896.
- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, 76(5), 820–830.
- Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78(1), 105–119.
- Wierzbicka, A. (1992). Talking about emotions: Semantics, culture, and cognition. *Cognition and Emotion*, 6, 285–319.

Affective bodies for affective interactions

Marco Vala, Ana Paiva and Mário Rui Gomes

1. Introduction

A decade ago it was nearly impossible to conceive of embodied characters directed by intelligent agents in a complex virtual world. Today, however, recent computer games and movies show us that embodied characters are far more than a possibility; they are already an active part of the public-at-large's imaginary worlds and real life.

However people demand more than merely animated characters. Embodied characters must be alive and convincing. They must act as a real actor would do and make the audience smile or cry. For this reason, in spite of the strong research and development we are witnessing in the area of synthetic characters, many new topics and interesting questions remain to be solved. How can we create believable characters for the virtual worlds of tomorrow? How can we offer richer interactions between the characters and their audience? How can we convey through the actions of the characters, some deeper internal state, making the characters more believable?

To address some of these questions, we maintain that embodied characters must be emotionally expressive (André *et al.* 2002). We have therefore proposed a method that modifies previously created animations for synthetic characters in order to generate affective movements in real-time for both humanoid and non-humanoid skeletons. The animator is free to model and animate the character without an imposed skeleton or any other design restrictions.

2. Related work

Amongst some of the most influential work in this area of synthetic characters, the toolkit Jack, developed by Badler's team, offers a generic way of modelling and animating humanoid characters (Badler *et al.* 1993). Similarly, Thalmann proposes human-like synthetic characters with detailed geometrical models and animation techniques (Fua *et al.* 1998; Kalra *et al.* 1998; Aubel & Thalmann 2000).

Both apply a “naturalistic” approach that puts emphasis on visual realism, which is not necessarily the same thing as believability. Rather than visually accurate movements, a character should have appropriate behaviour.

Some research groups consider autonomous control and action-selection mechanisms to trigger the most adequate action for each possible situation. Blumberg at the MIT Media Lab proposed several interactive creatures that behave autonomously (Blumberg & Galyean 1995; Russell & Blumberg 1999). Perlin developed responsive animated characters within the Improv system (Perlin & Goldberg 1996). Both argue that believability is the character’s ability to select the right behaviour to appear lifelike.

At the same time the scientific community has become increasingly aware of a new field of “affective computing” (Picard 1997; Tao *et al.* 2005; Paiva *et al.* 2007). We believe that displaying affective behaviour is not only the key to creating more believable characters, but also to achieving a new level of human-computer interaction (Paiva 2000; Pelachaud & Cañamero 2006). A particular task is to create characters that can adapt their current actions to denote emotions and inner feelings.

Perlin proposed Perlin Noise (Perlin 1995) as a way to add expressiveness to animated motions. The results are visually interesting particularly for idle movements. The random noise functions cannot however display a clear link to a specific emotion or inner feeling.

Unuma and colleagues (1995) presented a Fourier Functional Model, which describes human periodic motions and allows the generation of multiple walking and running cycles in real-time. This parametric approach would be extremely useful if it could be extrapolated for different movement classes. Unfortunately, the parameters provided by the functional model are very specific.

Amaya and colleagues (1996) use Emotional Transforms, which capture the difference between neutral and emotional motions. In particular, Amaya concluded that the speed and the amplitude of movements fluctuate between different emotions. Emotional Transforms had no further application although the results were in part confirmed by the EMOTE system.

EMOTE (Chi *et al.* 2000; Badler *et al.* 2000) is probably the most consistent and accurate model to date for creating expressive movements in real-time. It is able to dynamically change animations using the *effort* and *shape* parameters of Laban’s movement analysis. The current implementation, however, is not yet a full body approach (only arm and torso) and relies on the existence of particular bones in the skeleton (such as wrist or elbow). This requires specific parameters for each bone or class of bones that in turn can only work in very specific skeleton configurations.

That is perhaps the major limitation of all previous work. The single use of humanoid skeletons excludes non-humanoid creatures like animals or aliens. Our goal was to create a more general approach suitable for both humanoid and non-humanoid skeletons. Thus, and in the context of the SAFIRA project,¹ we have proposed affective bodies that are emotionally expressive (André *et al.* 2002). Our method modifies previously created animations in order to generate affective movements in real-time for both humanoid and non-humanoid skeletons. The animator is free to model and animate the character without an imposed skeleton or any other design restrictions.

3. Using the body to express emotions

Building affective bodies is a complex task. As Coulson (this volume) notes, people do not express or perceive emotions from body movements in the same way and universal recognition is unlikely (see the companion chapter *Expressing Emotion Through Body Movement*). This is in fact a major problem which prevents us from establishing direct mappings between a specific emotion and a set or subset of motion parameters. EMOTE for instance does not provide any mapping. The appropriate parameterisations should be provided by the character creator, taking into account the nature of the application and the cultural background of the target audience.

Our system generates affective animations in real-time using a set of neutral² animations and scripts with the right parameterisations for each available emotion. The character creator provides the neutral animations and the affective scripts. The system generates any combination of animation/emotion upon request. For example a neutral walk animation and a sad script would generate a sad walk animation.

Obviously the character must have a conceptual architecture that supports the system just described. The bottom level of the architecture manipulates the body of the character. Above this, a middle level manages the animations. The top level executes the affective scripts. This clear separation between geometry, animation and behaviour is inspired by the work of Blumberg (Russell & Blumberg 1999) and Perlin (Perlin & Goldberg 1996).

1. EU-funded project number IST-1999-11683 (<http://gaips.inesc-id.pt/safira>).

2. Neutral means without denoting any particular emotion.

3.1 A simple body structure

The structure that defines the body of the synthetic character plays a very important role in the whole system. The geometrical model must be simple enough to allow fast updates but complex enough to ensure good visual realism.

We use a *joint-link model*, as it is usually described in the literature, with a hierarchical skeleton and a deformable skin surface. The hierarchical skeleton is a collection of bones connected in a tree structure (Figure 1, left). The movement of a particular bone not only affects the bone itself but also its children, as in a real skeleton, where, when we move the elbow, the wrist and all the bones in the hand are also moved. The skin surface is one or more triangular meshes which delineate the shape of the synthetic character (Figure 1, right). Each skin point is connected to one or more bones, and the movement of the skeleton changes the skin surface accordingly.

The joint-link model guarantees a good trade-off between quality and low computational demand. It can cope with large detailed models and ensures a tractable number of calculations when updating the skin surface in real-time.

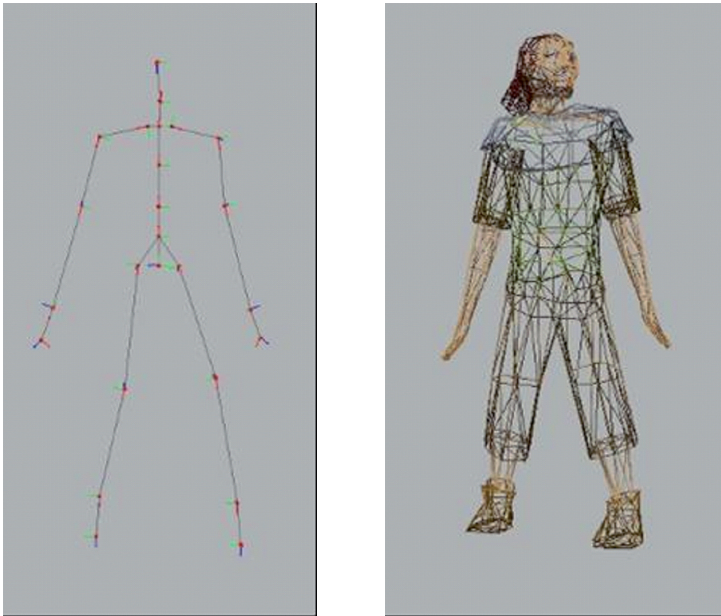


Figure 1. Hierarchical skeleton (left) and skin surface (right)

3.2 Key-frame animations

A character has a number of animations created offline. An animation is a key-frame sequence of skeleton posed over time. When an animation is played, the system updates the skeleton to reflect the current pose in the timeline. The skin surface is automatically adjusted using the internal mechanics of the joint-link model described previously.

There are three classes of animations: stances or poses (animations with a single key-frame), actions (animations with a sequence of key-frames that runs once and stops), and loops (animations that run in an endless cycle). All the animations must maintain a relative bone position and orientation towards the initial pose of the character. This is an important pre-requisite of the system and the original animations must be converted if they have absolute bone values. Below are the equations that perform the conversion.

The position of each bone in the bone space (bt_m) is the difference between its position in the world space (wt_m) and the position of the parent in the world space (wt_n) transformed by the inverse orientation of the parent in the world space (wr_n^{-1}); the difference for the neutral pose is calculated by subtracting the neutral position (nbt_m):

$$bt_m = [wr_n^{-1} \times (wt_m - wt_n)] - nbt_m \quad m, n = 0 \dots bone \quad n = parent(m) \quad (1)$$

The orientation of each bone in the bone space (br_m) is the orientation in the world space (wr_m) after transformed by the inverse orientation of the parent in the world space (wr_n^{-1}), and by the inverse neutral orientation (nbr_m^{-1}):

$$br_m = nbr_m^{-1} \times (wr_n^{-1} \times wr_m) \quad m, n = 0 \dots bone \quad n = parent(m) \quad (2)$$

3.3 Stance composition and motion parameters

Once all the animations are in the bone space, the system is ready to play the animations and process the affective scripts. The affective scripts can modify ongoing animations using a couple of procedures, namely the stance composition and changes to some motion parameters.

The first procedure comes from an empirical idea, often used in cartoons, which identifies an emotion with a particular pose. For example, when a cartoon is *sad*, it usually walks bending the torso and facing the ground. We can use these “affective poses” and combine them with neutral animations to generate affective animations in real time (Figure 2).

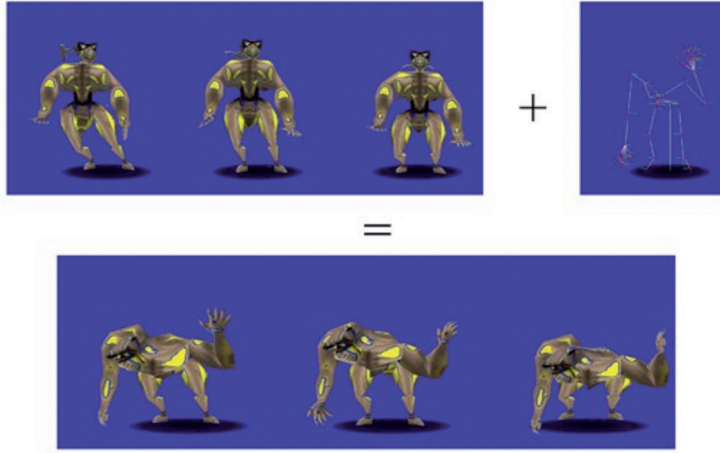


Figure 2. Stance composition: neutral movement + sad stance = sad movement

Internally, the system performs a spherical linear interpolation between each bone in the stance and the matching bones in the current frame of the neutral animation. As a direct result the stance will influence and drive the overall movement of the synthetic character. Below are the equations that perform the stance composition.

The combined position of each bone (bt_m) is the sum of the position in the current frame (fbt_m) with the weighted position in the active stance (sbt_m):

$$bt_m = fbt_m + (sbt_m \times \text{weight}) \quad m = 0 \dots \text{bone} \quad (3)$$

The combined orientation of each bone (br_m) is the weighted orientation in the active stance (sbr_m) transformed by the orientation in the current frame (fbr_m). In this case, the weight influence is calculated using a spherical linear interpolation (*slerp*) between the identity rotation (I) and the orientation in the active stance:

$$br_m = fbr_m \times \text{slerp}(I, sbr_m, \text{weight}) \quad m = 0 \dots \text{bone} \quad (4)$$

Notice that the spherical linear interpolation uses a weight value which defines how much the stance influences the neutral movement. Therefore, the stance composition can be more or less exaggerated to reflect different emotional intensities. Following the example of the sad character, the torso could be more or less bent thus giving an idea of how sad the character is.

The second procedure extends the results of Emotional Transforms (Amaya *et al.* 1996). Amaya reported that the speed and spatial amplitude of movement vary noticeably with different emotions. We use these two motion parameters to embody or emphasize particular emotions. For example, happy movements are

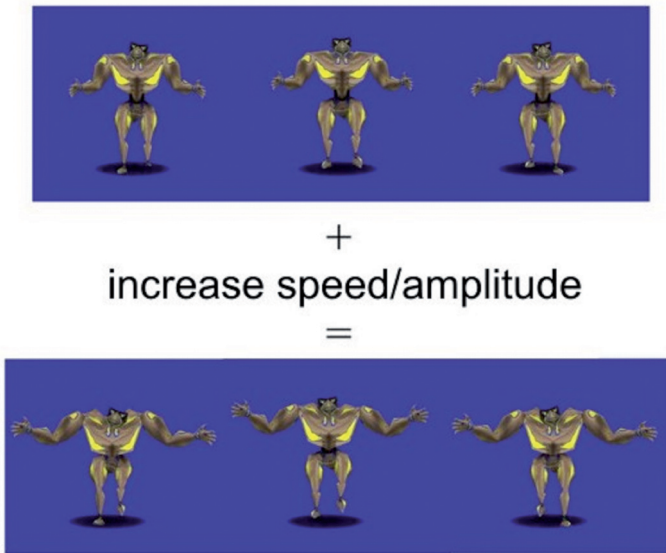


Figure 3. Speed/Amplitude variation: neutral walk + increase = happy walk

usually fast and wide; we can change a neutral walk and increase the speed/amplitude to generate a happier walk (Figure 3).

The speed parameter is directly related to the animation frame rate. Increasing the speed of an animation reduces the time between two consecutive frames, thus making the execution time faster. In the same way, decreasing the speed of the animation increases the time between two consecutive frames, thus making the execution time slower. Note that increasing the time between frames might affect the overall quality of the animation because it significantly reduces the number of frames per second. In order to preserve the smoothness of the movement the system is able to interpolate between frames using a spherical linear interpolation between two consecutive frames. The spatial amplitude parameter defines how wide/narrow the animation is. Amaya concludes that some emotions generate movements with a wider signature than others. Using the cartoon example once more, angry characters usually spread their arms to intimidate while scared ones assume a more defensive and contracted attitude. Increasing the spatial amplitude increases the angles between the bones of the skeleton along the animations, thus creating wider movements. Decreasing the spatial amplitude of the animation reduces the angles between the bones of the skeleton along the animations, thus creating narrower movements.

The system uses a weight function that represents the amplitude. The weighted position of each bone (bt_m) is equal to the current position (bt_m) multiplied by the *weight*:

$$bt_m = bt_m \times weight \quad m = 0 \dots bone \quad (5)$$

The weighted orientation of each bone (br_m) is the result of the spherical linear interpolation (*slerp*) between the identity rotation (I) and the current orientation (br_m) using the *weight* as balance factor:

$$br_m = slerp(I, br_m, weight) \quad m = 0 \dots bone \quad (6)$$

3.4 A minimal resource mechanism

The playback of animations and the execution of affective scripts can generate potential conflicts between different animations. Our system follows a small set of rules which regulate the coexistence of animations being played at the same time:

- The animation engine cannot play more than one loop or action at the same time, but the stances can coexist with both of them;
- The loops are played continuously until the animation engine receives a request for another loop or action, or an explicit command to end the loop;
- The actions are always played until the end, and cannot be interrupted by any request;
- The stances are combined in real-time with the active loop or action, but the animation engine cannot handle more than one stance at the same time.

Clearly this is a minimal resource mechanism that prevents two loops or actions from being played at the same time and offers the necessary flexibility to introduce stances in the middle of the animation in play (allowing the stance composition that was introduced before). These rules guarantee that all the animations run properly without ambiguous visual results.

Note that although loops and actions cannot occur at the same time the system can blend them in a smooth manner to ensure a visually accurate transition.

3.5 The animation cycle

The current implementation of the animation cycle is very straightforward and does not offer anything new. It receives requests for animations (loops, actions or stances), and plays them sequentially. If the resource mechanism that manages the coexistence of animations blocks a particular request, then this request is

queued until it can be played. When there are no more requests and the character is idle, the animation engine automatically plays an idle behaviour. At the end of the cycle, the animation engine updates the body of the synthetic character so it can reflect the active animations during the previous round.

4. Creating affective scripts

The definition of an affective script is very similar to the act of creating an animation for a particular character. The designer/animator configures a certain combination of stances and motion parameters that will produce the desired affective result. We do not claim that a certain combination of stances and parameters intended to denote *sadness* always produces a visual result that everyone can call “sadness”. As we have mentioned before, we are still very far from such a universal recognition. But we do claim that is easier and faster to play a little bit with these parameters (to create the visual impression of sadness) than to model the same behaviour as single animations. Moreover a single script can be used to influence several neutral animations and otherwise the animator would have to create a new emotional version for each animation.

Next are some examples of scripts using the characters of FantasyA (the next section presents a brief introduction to FantasyA). These examples are purely empirical and do not have any emotional theory behind them. Figure 4 shows Nakk performing a walking loop. A *gloat* script containing a gloat stance modified the neutral walking animation and the character assumed a gloating attitude by raising the arms in a defiant manner.

Figure 5 exemplifies a *sadness* idle-behaviour. Ronin’s idle cycle was modified by a sadness script which uses a stance with the character is looking down and decreases the speed and the spatial amplitude to an half.

Figure 6 shows Alveggha with *fear*. Alveggha’s idle cycle was modified by a fear script which uses a stance where the character is using the hands to protect the face. The result is an animation where the character seems to be afraid of something.



Figure 4. Nakk gloating while walking (walk motion + “gloat” stance)



Figure 5. Ronin sad (idle motion + “sad” stance, speed = 0.5, spatial-amplitude = 0.5)



Figure 6. Alvegha with fear (idle motion + “fear” stance)



Figure 7. Feronia angry (idle motion + “angry” pose, speed = 1.8)

The last example in Figure 7 represents an *angry* mood. Feronya's idle cycle was modified by an angry script with an arguing pose and a sudden increase of speed. The result shows the character gesticulating as if she was angry with something.

5. Applications

The first prototype of the system was used in FantasyA (Arafa *et al.* 2002; Höök *et al.* 2003), one of the demonstrators within SAFIRA. FantasyA is a magic duel where two characters fight each other using spells (Figure 8). However, unlike traditional computer games, the player does not decide which magic spell to use. Instead s/he influences the emotional state of character and it chooses the spells autonomously. Human players control their avatar using a sympathetic interface, SenToy (Figure 9), that is able to recognize six emotions: sad, happy, gloat, angry, fear, and surprise. At each turn, the combination of the emotional states of the two characters (the one controlled by the player and the computer opponent) results in spells, either offensive or defensive, which can damage the opponent or protect the character from future attacks. For example, if the computer opponent is gloating and the player uses the SenToy to influence his avatar to become angry, that will surely lead to an offensive action, probably a blast. Then, according to the results of the action, there is a reaction phase where both characters change their emotional state in response. Following the previous example, if the blast succeeds then the computer opponent becomes fearful and the character controlled by the



Figure 8. Ongoing duel in FantasyA

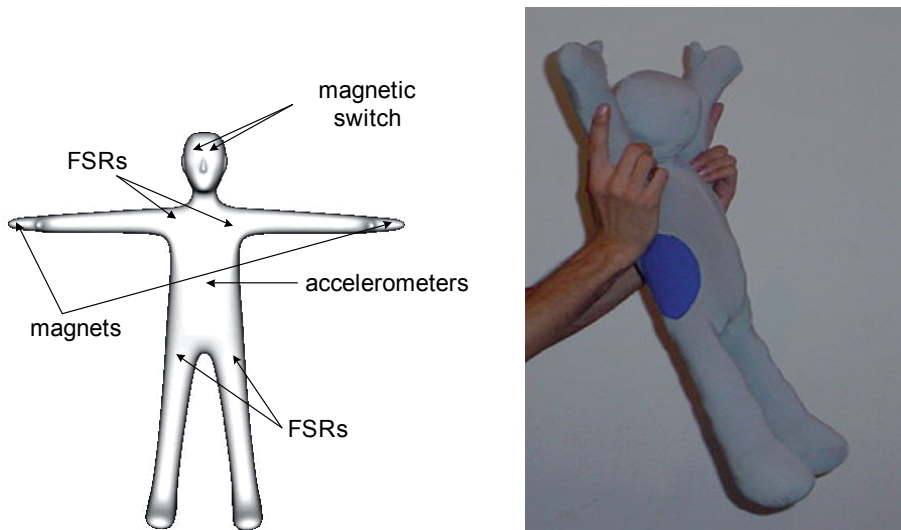


Figure 9. SenToy

player becomes happy. The game proceeds with the opponent's turn and so forth, until the end of the duel.

To effectively play the game, players must understand the affective display. The player must be able to recognize all the emotional states in order to discover the combinations that lead to each spell. Mastering the affective interactions is the only way to win the game.

6. Results

We presented the first prototype of our system and a single case of its use. FantasyA's characters demonstrate that we can successfully generate affective animations in real-time. The preliminary results borrowed from the evaluation of FantasyA demonstrate the creation of a good level of empathy between the characters and the audience (Anderssen *et al.* 2002). The affective bodies are in our opinion a step towards more believable synthetic characters.

Obviously our approach still has some limitations. The resource mechanism is too strict and as an example the character cannot scratch the head while walking because actions cannot coexist with loops. Clearly a walking movement uses primarily the legs and a scratching movement uses the arm; there is no conflict at all. A more advanced resource mechanism at the bone level would allow both

animations. Note that the current resource mechanism is a minimalist solution which will surely be improved in the future.

Another flaw of the system is the possibility of self-intersections or violation of physically valid poses as the result of the “blind” composition of animations. But this is the inevitable trade-off with a generic composition algorithm that handles all bones in the same way. We cannot apply different algorithms and use different parameters for the arm and for the wrist; that would ruin the advantage of a generic approach for both humanoid and non-humanoid skeletons. We think that some physical constraints at the bone level might overcome this weakness.

7. Conclusions

We have presented a system for generating affective animations in real-time. Our approach uses neutral animations together with stance compositions and/or changes in the speed/amplitude of the animation. The major contribution of this work is the possibility of modifying ongoing animations using parameters to create affective movements rather than having to model each pair emotion/animation offline. The method can also be used in completely different characters separately from the internal skeleton configuration or animation technique; it is valid for both humanoid and non-humanoid skeletons.

The architecture behind a character is based on a behavioural approach. We argue that if we aim at believable characters, the ability to convey a certain inner-life is more important than visual realism. We think that displaying affective behaviour is the next step towards a new level of believability.

Acknowledgements

This work has been partially supported by the EU-funded SAFIRA project, number IST-1999-11683.

References

- Amaya, K., Bruderlin, A. & Calvert, T. (1996). Emotion from Motion. In R. Bartels & W. A. Davies (Eds.), *Proc. Conference on Graphics Interface '96* (pp. 222–229). Toronto, Ontario: Canadian Information Processing Society.
- Anderssen, G., André, E., Arafa, Y., Botelho, L., Bullock, A., Chaves, R., Fleischmann, M., Figueiredo, P., Gebhard, P., Gonçalves, B., Goulev, P., Höök, K., Li, Y., Liesendahl, R., Mar-

- tinho, C., Paiva, A., Petta, P., Ramos, P., Sengers, P., Strauss, W., Vala, M. & Wolf, M. (2002). *SAFIRA Deliverable 7.3 – Final Evaluation Report*.
- André, E., Arafa, Y., Fleischmann, M., Gebhard, P., Geng, W., Kulesa, T., Paiva, A., Sengers, P., Strauss, W. & Vala, M. (2002). *SAFIRA Deliverable 5.2 – Shell for Emotional Expression*.
- Arafa, Y., Bullock, A., Chaves, R., Costa, M., Fleischmann, M., Goulev, P., Höök, K., Liesendahl, R., Magar, W., Martinho, C., Paiva, A., Peixoto, M., Piedade, M., Prada, R., Rebelo, F., Sengers, P., Strauss, W., Sobral, D., Vala, A. & Vala, M. (2002). *SAFIRA Deliverable 6.2 – Final Prototypes of the Demonstrators*.
- Aubel, A. & Thalmann, D. (2000). Realist Deformation of Human Body Shapes. In N. Magnenat-Thalmann, D. Thalmann, & B. Arnaldi (Eds.), *Proc. of Computer Animation and Simulation 2000: Proceedings of the Eurographics Workshop* (pp. 125–135). Wien New York: Springer Computer Graphics.
- Badler, N., Philips, C. & Webber, B. (1993). *Simulating Humans: Computer Graphics Animation and Control*. New York, NY: Oxford University Press.
- Badler, N., Costa, M., Zhao, L. & Chi, D. (2000). To Gesture or Not to Gesture: What is the Question? In *Proceedings of Computer Graphics International 2000*, Geneva, Switzerland, June 19–24, 2000 (pp. 3–10). IEEE Press.
- Blumberg, B. & Galyean, T. (1995). Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. In *Proceedings of SIGGRAPH'95* (pp. 47–54).
- Chi, D., Costa, M., Zhao, L. & Badler, N. (2000). The EMOTE Model for Effort and Shape. In *Proceedings of SIGGRAPH'2000* (pp. 173–182). New Orleans.
- Fua, P., Plänkers, R. & Thalmann, D. (1998). Realistic Human Body Modeling. In *Proc. Fifth International Symposium on the 3D Analysis of Human Movement*, Chattanooga, TN, July 1998.
- Höök, K., Bullock, A., Paiva, A., Vala, M., Chaves, R. & Prada, R. (2003). FantasyA and SenToy. In *Proc. Conference on Human Factors in Computing Systems – CHI '03 extended abstracts on Human factors in computing systems* (pp. 804–805). New York, NY: ACM Press.
- Kalra, P., Magnenat-Thalmann, N., Moccozet, L., Sannier, G., Aubel, A. & Thalmann, D. (1998). Real-Time Animation of Realistic Virtual Humans. *IEEE Computer Graphics and Applications*, 18(5), 42–56.
- Paiva, A. (Ed.) (2000). *Affect in Interactions: Towards a new generation of interfaces*. Heidelberg & Berlin: Springer-Verlag.
- Paiva, A., Prada, R. & Picard, R. (Eds.) (2007). *Affective Computing and Intelligent Interaction, Second International Conference, ACII 2007*, LNCS 4738. Berlin/Heidelberg: Springer.
- Pelachaud, C. & Cañamero, L. (Eds.) (2006). Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids. Special issue of the *International Journal of Humanoid Robotics*, 3(3).
- Perlin, K. (1995). Real-time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics*, 1(1), 5–15.
- Perlin, K. & Goldberg, A. (1996). Improv: A System for Scripting Interactive Actors in Virtual Worlds. In *Proceedings of SIGGRAPH'96* (pp. 205–216). New York, NY: ACM Press.
- Picard, R. (1997). *Affective Computing*. Cambridge, MA: The MIT Press.
- Russell, K. & Blumberg, B. (1999). Behavior-Friendly Graphics. In *Proceedings of Computer Graphics International 1999* (pp. 44–50, 241), Canmore, Alta., Canada. June 7–11, 1999. IEEE Press.
- Tao, J., Tan, T. & Picard, R. (Eds.) (2005). *Affective Computing and Intelligent Interaction, First International Conference, ACII 2005*, LNCS 3784. Berlin/Heidelberg: Springer.

- Unuma, M., Anjyo, K. & Takeuchi, R. (1995). Fourier Principles for Emotion-based Human Figure Animation. In *Proceedings of SIGGRAPH'95* (pp. 91–96). New York, NY: ACM Press.
- Vala, M., Paiva, A. & Gomes, M. R. (2002). From Virtual Bodies to Believable Characters. *AISB Journal*, 1(2), 219–223.
- Vala, M. (2003). *From Virtual Bodies to Believable Characters – Reusable Synthetic Characters with Expressive Bodily Behaviour*. Unpublished MSc Thesis, IST, Technical University of Lisbon, Portugal.

Animating affective robots for social interaction

Lola Cañamero

1. Introduction

The social and the emotional are highly intertwined. For some researchers, emotions come into play (primarily for some, exclusively for others) as soon as we consider individuals in interaction with their social environment. For others, emotions are at the very heart of what being social means. Some opinions even establish a relation of “identity” between these notions, somewhat like the two sides of a coin (e.g., Dumouchel 1999; also this volume). In humans, emotions influence and shape the development of sociality as much as sociality influences and shapes the development of emotions (see e.g., Lewis & Granic 2000; Nadel & Muir 2005).

This strong interdependence is being increasingly echoed in AI research. In this respect, the area of socially intelligent agents (Dautenhahn 1998; Dautenhahn et al. 2002) has witnessed over the last years a growing interest in the involvement of emotions in social interactions. Likewise, the affective computing community has devoted considerable effort to the design and implementation of models of emotions for social interaction (see, for example, Trappl & Petta 1997; Cañamero 2001b; Paiva 2001; Aylett & Cañamero 2002; Breazeal 2002; Prendinger & Ishizuka 2004; Tao *et al.* 2005; Pelachaud & Cañamero 2006; Paiva *et al.* 2007).

Robots are not humans, though, and one can always question whether social interactions involving artifacts need to take emotional aspects into consideration – it certainly makes sense to ask “Who needs emotions?” as Fellous and Arbib (2005) put it. In previous papers, I have discussed some of the benefits of integrating emotion-based mechanisms in the control architectures of adaptive autonomous agents, in particular robots, from the perspectives of their design (Cañamero 2001c, 2003), and of their contributions to multidisciplinary research towards emotion understanding (Cañamero 2005; Cañamero & Gaussier 2005). In this chapter, I discuss the value of some of those mechanisms from the standpoint of social interaction – in particular, their relevance for animating robots

“from the inside out,” and their relation to (human-oriented) emotional expression and perception. I will thus consider only interactions involving robots and humans, although the investigation of such mechanisms and their expression in purely artificial societies can also pose many interesting research questions and shed light into the evolution and uses of affect-related expression used as reference and signaling mechanisms in social interaction (see e.g., Lowe *et al.* 2004, 2005). In doing so, this chapter takes the reader through some personal thoughts on some fundamental questions such as: Are observable behavioral features enough, or is an underlying model for emotion synthesis needed? In this latter case, how can we ground emotions in the architecture of the robot so as to give rise to meaningful interactions? Can we take inspiration from research on human emotions for this?

2. Emotions and social robots

What can artificial emotions contribute to social interactions between robots and humans? As pointed out in (Cañamero 2001a; Cañamero & Gaussier 2005), we can intuitively think of different roles that emotions can play in social interactions between robots – and more generally artifacts – and humans:

Conveying intentionality

People need to understand the behavior of others as resulting from causes or intentions that allow them to form coherent explanations of their observations. This coherence is necessary to interpret past or current relations, make predictions and establish expectations about future behavior. Emotions and personalities are often postulated as causes of behavior and as sources of intentions when explaining the behavior of other humans and animals, and it is easily extrapolated to objects, particularly to (interactive) technology (Reeves & Nass 1996). Autonomous robots can, in addition, use emotions and their expressions to convey intentions or needs to humans, and in some cases have limited recognition of, and response to, affective states of humans, therefore closing the interaction loop.

Eliciting emotions

In the same way as other people's emotions elicit emotional responses from humans, robots' emotions can be used with the same purpose, seeking responses that either match the robot's emotional state or are instrumental to it – e.g., a robot unable to accomplish a task can use some form of expression of distress to seek help from a human “caregiver”.

Human comfort and acceptance

A fundamental concern is whether humans are willing to accept and trust robots as social partners, and how they should be designed to favor acceptance by humans. Two key factors for this seem to be believability, and that the human feels that s/he is (and that s/he really is) in control of the interaction at critical points. Robots able to express emotions and adapt their interactions to the emotional state of their partners can make humans feel more comfortable during interaction. One obvious reason is that this interaction is tailored to meet the emotional needs of the human. Another important reason is that coherent emotional behavior and expressions make the robot more believable (Ortony 2003), as in a sense it is perceived as “closer” or more similar to ourselves (or generally to a living being). These factors depend on many other variables, such as the personality of the human partner, the application domain where the robot is put to work, the cultural and social context, etc. Designing “generic” emotional robots embedding general emotion models that do not take into account individual differences seems thus to have clear limitations. It would be more advisable (although much more difficult) to model emotions adapting them to types of “profiles” that take all these factors into account, and also reflect the evolution of emotional interactions over time – i.e., their “affective history” (Blanchard & Cañamero 2005).

Enhanced communication

Emotional expression is a key element in non-verbal communication, as endowing a robot with emotional expressions can make communication cognitively less expensive for the human partner. If emotional artifacts are to achieve a sufficient level of sophistication at some point in the future, we could also dream of communicating with them at a “deeper” level of understanding. For example, provided that one day they could be able to interpret our subtle expressions and obtain relevant information from contextual clues, we would also like that they “understand” what we *mean*, not (only) what we say.

3. Surface or beyond?

... the concepts discussed in this article characterize the psychological basis of believability. Storytelling, empathy, historical grounding (autobiography), and “ecological grounding” in an environment are identified as factors relevant to the way humans understand the (social) world. [...] It is hoped that approaches to achieving believability based on these concepts can avoid the “shallowness” and “cheating” of approaches to believability that merely take advantage of the anthropomorphizing tendency in humans. (Dautenhahn 1998, p. 574)

The same division of opinions that Dautenhahn portrays in the area of socially intelligent agents can be seen in the emotion modeling community: Should we model (only) the observable, surface aspects of emotions, or must we go for a “deep” modeling of a “true” emotional system beyond surface? Which is possible, necessary, and/or sufficient?

Whereas emotion modeling (synthesis) for individual agents has generally focused on the design of architectures to endow these agents with “true” emotional systems, the design of emotions for agents interacting socially has primarily paid attention to the external or “superficial” features of emotional expression and interaction. Believability being a major issue as soon as humans are involved in (social) interactions with robots, the question can be rephrased as: What makes an emotional robot believable for social interactions?

The features that allow humans to *perceive* the emotional displays of the robot as believable enough to elicit appropriate responses from them seem to be the most apparent answer. One would immediately think of general expressive features, such as something resembling a face (human or not) with some typical elements like (moving) eyes, eyebrows, or a mouth; patterns of movement; posture; timely responses; inflection of speech; etc. Other expressive features more related to social interaction include turn-taking during the interactions; direction of gaze; eye-to-eye contact; joint attention; etc. In theory, all these features could be modeled taking a “shallow” approach – which does not make their implementation more easy or trivial. However, such an approach makes it very difficult to maintain believability over prolonged interactions, as it is unlikely that the *behavior* of the robot will remain coherent over sustained periods of time, given the complexity of these interactions. Coherence of behavior is an important issue (Cañamero 2001a; Ortony 2003), as humans want to understand and explain observed (expressive) behavior as the result of some underlying causality or intentionality for the emotional interaction with the robot to be believable and acceptable to them. This can only be properly achieved if expressive behavior is guided by some underlying system of emotion synthesis (and ideally of personality as well).

I adhere thus to the opinion that the believability of emotional displays and interactions can be better achieved, for non trivial social interactions, if it is sustained by a “deeper” emotion system grounded in the architecture of the artifact. This is not to say, however, that I consider “shallow” approaches as uninteresting or “cheating”. The fact that the human tendency to anthropomorphize is so pervasive and compelling that makes us treat our TV set and computer like people (Reeves & Nass 1996) makes it worth studying it with the possibilities that emotional artifacts offer. Much can be learned about human emotions and emotional interactions from projects that heavily rely on the human tendency to anthropomorphize, such as the expressive robots Sparky (Scheeff *et al.* 2002), Kismet

(Beazeal 2002) or Felix (Cañamero & Fredslund 2001; Cañamero 2002). On the human side, they can help us identify key elements that shed light on what triggers in humans the tendency to anthropomorphize and what makes emotional behavior and displays believable to the human eye. On the robot side, they can provide very valuable feedback for the design of robots (e.g., their morphology) that can interact socially and emotionally in a way that is more adapted and natural to humans.

The choice between one approach and the other will thus depend on what we are interested in learning about emotions, and on the application foreseen for the robot.

4. Anchoring emotions in robots

If we want our emotional robots to go beyond superficial emotional features and displays to “have” emotions (in some sense) and deeper emotional interactions, we must find ways to anchor emotions in them that are suited to the robot’s structure and dynamics of interactions, as well as to the other partners (including humans) in their social world. This anchorage can be understood in a weaker or a stronger sense.

In its weaker sense, it can be taken to mean endowing robots with some components or modules that explicitly represent some elements of emotions. This set of components produces, given appropriate inputs, outputs (e.g., behavior, textual or graphical displays) appearing to arise from emotions because they are similar to emotional behaviors and responses observed in biological systems under equivalent circumstances. This corresponds to the “black-box” approach to emotion modeling in the classification proposed by Wehrle and Scherer (1995) and discussed in (Cañamero 2003) in the context of emotions for action selection in autonomous agents. As pointed out by Wehrle:

although such models provide little information concerning the mechanisms involved, they are very useful for practical decision-making and for providing a sound grounding for theoretical and empirical studies. In particular, they can help to investigate the necessary and sufficient variables. System performance (e.g., the quality of classification and computational economy), as well as cost of data gathering, are important criteria for assessing the quality of the chosen computational model. (Wehrle 2001, p. 565)

This form of anchoring emotions can thus be very valuable to solve AI and robotics problems such as speeding up the system’s responses to certain stimuli. It can also be of great help to do a systematic analysis of significant variables. However,

the emotional system of the artifact does not, *per se*, shed much light on the underlying emotional mechanisms and processes, nor is it fully meaningful to the robot itself, as it has been engineered by the designer. Only in a very restricted or metaphorical sense would I say that this form of anchorage allows the robot to “have” emotions.

In its stronger sense, the anchorage of emotions can be seen as emotion grounding, in the same sense of the term “grounding” as it is given when talking about the *grounding problem* (Harnad 1990). In this sense, emotions must be modeled in such a way as to be rooted in, and intertwined with, the perception and action processes of the robot so that emotions and their consequences can have an intrinsic meaning for it. To put it in Wehrle’s words:

grounding somehow implies that we allow the robot to establish its own emotional categorization which refers to its own physical properties, the task, properties of the environment, and the ongoing interaction with its environment.

(Wehrle 2001, p. 576)

Process modeling (Wehrle & Scherer 1995), which attempts to simulate naturally occurring processes using hypothesized underlying mechanisms, is a more appropriate approach than the black-box one in this case. In this stronger sense, emotions being intrinsically meaningful to the artifact itself, I would argue that the artifact “has” emotions in a broad sense that implies the adoption of a functional view on emotion modeling (Frijda 1995; Cañamero 1998; Cañamero & Avila-García 2007). This position is not devoid of methodological problems, though, as discussed in (Cañamero 1998; Wehrle 2001). For example, if we take inspiration from models of existing (biological) emotional systems (e.g., emotion categories, dimensions, action tendencies) to design the robot’s emotional system, one can question to what extent its emotions are really grounded. On the contrary, if we let the robot develop its own emotions, these might not be understandable to us. As a partial way out of this dilemma, Wehrle proposes the use of appraisal dimensions borrowed from psychology as a basis for the value system of the artifact, in order to benefit from the possibility of describing the resulting emotional behavior in known terms.

In the following sections, I will sketch some elements of an alternative view to grounding emotions in embodied artifacts, in line with the emotion architectures proposed in (Cañamero 1997; Cañamero & Avila-García 2007). I will first discuss some elements necessary to ground emotions in individual artifacts, to consider later additional elements to ground emotions in artifacts interacting socially.

5. Grounding emotions in the individual

Our interest in emotion in the context of AI is not an interest in questions such as ‘Can computers feel?’ or ‘Can computers have emotions?’ [...] our view is that the *subjective experience* of emotion is central, and we do not consider it possible for computers to experience anything until and unless they are conscious. Our suspicion is that machines are simply not the kinds of things that can be conscious. (Ortony *et al.* 1988, p. 182)

Although the use of terms such as “vision” or “memory” seems to be generally well accepted when applied to artifacts, in spite of the fundamental differences between human and artificial vision or memory systems, many researchers of (artificial and human) emotions remain skeptical about the use of the term “emotion” applied to artifacts and about the possibility for robots to “have” emotions and emotional interactions with humans and other agents. Arguments commonly heard stress the impossibility for robots to implement notions such as self, feelings, subjective experience, or consciousness. These phenomenological aspects of experience and emotions are most apparent to (“typical”) humans, and we seem to be particularly keen on using them to define the realm of what is uniquely human.

I share with these criticisms the skepticism about the possibility for robots and other artifacts to have selves, feelings, subjective experience, consciousness, and emotions *in the same way as humans do*. Robots are not biological systems, they are made from a totally different sort of material, have different bodies, actuators, perceptual and cognitive capabilities, experiences, and niches, and in a sense they can be regarded as a different, “new” type of species. However, I do believe that endowing robots with some form of (at least some of) these notions – or rather their functional counterparts – is necessary for them to interact with humans in a way that humans can understand and accept, and can likewise enhance many aspects of their behavior and performance. Dogs do not have human emotions either, but they certainly engage in social and emotional interactions with us, and vice versa. Rudimentary implementations of some of these concepts that ground emotions have already been proposed by practitioners of the embodied AI and robotics approach, as we will see below. We must keep in mind, however, that these are only the first attempts of a nascent endeavor, and that at this stage we can only aim at laying foundations rather than at witnessing accomplished results.

As already mentioned in the previous section, this “strong” view raises the question whether (and why) we are willing to/should use terms borrowed from models of human cognition and emotion applied to these artificial systems. Arguments can be set forth against and in favor of this practice. The use of these terms rather than newly invented ones lets *us*, humans, explain behavior and phenomena in terms we already know and understand. However, I also think that great

care has to be paid to make this fundamental difference between artificial and human emotions (and the other notions as well) very clear, in particular when presenting our work to the general public, to avoid the dangers of over-attribution and frustrated expectations. These dangers might make it advisable to avoid the use of these terms in some contexts. Perhaps with time and habit one day people will talk about “robotic selves” and “emotions” as naturally as they already do about “computer vision”.

Below, I will consider the particular case of an adaptive embodied autonomous artifact – a robot – but many of the arguments can be easily applied to other artifacts such as software agents or virtual characters. I will only sketch some ideas around the notion of “self,” and deliberately avoid the more slippery grounds of feelings and consciousness. I think that, for these latter notions, it would be too premature to even give an opinion as to whether we will ever be able to implement (a rudimentary notion of) them in the future, given on the one hand to the lack of agreement and partial knowledge that the different disciplines have today, and on the other hand to the current state-of-the-art in AI and robotics, still at a too early stage for this. I refer the reader to (Damasio 1999) for his excellent account of “the feeling of what happens” (i.e., the notions of feelings and consciousness grounded in the body) from a neurobiological perspective, also full of insights for the design of artificial systems.

5.1 Embodiment

For the argument’s sake, I will begin by placing myself within an embodied AI perspective (Brooks 1991) and claim that this approach is better suited than symbolic AI to ground emotions in artifacts. In embodied AI, “embodiment” does not only mean that the artifact has a physical body through which it senses and acts on the world. As it is apparent to the practitioners of this field, embodiment has the fundamental implication that intelligence – cognition and, let us add, emotion – can only develop through the interactions of an embodied nervous system (or, for that matter, brain or mind) with its physical and social world. The emphasis of situated AI on complete creatures in closed-loop bodily interaction with their (physical and social) environment allows for a more natural and coherent integration of emotions (at least the “non-cognitive” or perhaps the “non-conscious” aspects of them) within the global architecture and behavior of the agent. This view has important implications for emotion grounding:

Closed loop

Emotion grounding requires that our model of emotions clearly establish a link between emotions, motivation, behavior, and perception, and how they feed back

into each other. This link makes that emotion (as well as motivation, behavior, and perception) can affect and be affected by the other elements in a way that can be either beneficial – e.g., energize the body to allow the individual to escape faster from a predator – or noxious – e.g., cause perceptual troubles that lead to inappropriate behavior – for the agent.

Bodily interaction

This link must be grounded in the body of the agent – for instance, by means of a synthetic physiology as in (Cañamero 1997) – since it is through the body that agents interact with the physical and social world. I am thus implying that emotions, natural or artificial, cannot exist without a body – although this is not the case in programs or agents that reason about emotions.

Value systems

Since we are talking about complete autonomous robots, emotions must be an integral part of their architecture. This means that emotions must be grounded in an internal value system that is meaningful (adaptive) for the robot's physical and social niche. It is this internal value system that is at the heart of the creature's autonomy and produces the valenced reactions that characterize emotions.

5.2 Robotic “selves”

Our subjective experience or the notion of “(our-)self” is the result of many different factors involving higher- and lower-level mechanisms and processes. Let us consider here two of these elements that have already received some attention in robotics research.

Bodily self

The experience of the own body is perhaps the most primary form of a notion of “self” and the starting point to endow an embodied artifact with some rudiments of this notion. Oliver Sacks eloquently illustrates how the perception of the own body grounds the experience of the self in his account of the case of the disembodied lady (Sacks 1986). The sense of the body is given, following Sacks, by various mechanisms working together that give rise to different body models in the brain:

- Vestibular feedback, which provides us with a sense of balance.
- Exteroceptive feedback such as visual feedback that gives rise to body-image (the brain's visual model of the body), and auditory feedback.
- Proprioception: the perception of the elements of our body (limbs, facial muscles, etc.) that makes us feel our body as belonging to us.

If one of them fails, the others can compensate to some extent. Due to a sensory polyneuritis, Christina, aged 27, lost the perception of her own body (proprioception), feeling “disembodied” or with a “blind body”. As a consequence, she initially lost the capacity to walk or move her limbs; her voice became flat, as vocal tone and posture are proprioceptively controlled; her face became also flat and expressionless (although her emotions remained of full and normal intensity), due to the lack of proprioceptive tone and posture; and she lost her sense of corporeal identity, leaving her with the impression that she could not “feel”. With strong self-motivation and rehabilitative support, she developed compensatory forms of feedback – very artificial-looking at the beginning, more natural with time and practice – that allowed her to become operational again; for example, she used attention and visual feedback to move her limbs, and learned to talk and move as if she was on stage. However, she could never recover the sense of bodily identity.

The elements that inform the sense of the body – vestibular system, visual feedback and proprioception – have been implemented in one form or another in different robotics projects, like for example in the humanoid robot Cog (Brooks *et al.* 1998). Proprioceptive feedback, with applications in robotics such as detecting self-position or controlling self-movement (including expression), has for example being used in the humanoid robot Infanoid (Kozima 2002) in conjunction with a value system to evaluate (in a valenced way) the proprioceptive information that the robot is receiving. Although still at its earliest stages, robotics research is thus starting to implement some rudimentary elements involved in the notion of bodily self. However, how these elements work together and interact with other subsystems to give rise to the *sense* of bodily identity belongs more to the realm of feelings and subjective experience, and therefore is out of our possibilities given the current state of the art in AI and robotics, and the still very partial picture that the different disciplines can offer about these notions. Our best robots today look more like Christina in the process of recovering from the effects of her sensory neuritis than like healthy human beings.

Autobiographic self

Another key element defining the notion of self is the “autobiographic self,” which Damasio (1999) presents as a necessary ingredient of the notions of identity and personality. Following Damasio, the autobiographic self is formed by the reactivation and coherent organization of selected subsets of autobiographic memories – past experiences of the individual organism. These memories are not static, but modified with experience during the individual’s lifetime, and are also affected by expectations about the future. The autobiographic self and autobiographic memories are deeply related to emotions in several ways. First, some of these autobiographic memories are emotionally loaded. Second, emotions facilitate mood-

congruent recall of past memories (Bower 1981). Third, as Damasio points out, the existence of an autobiographic memory and self allows organisms to provide generally coherent emotional and intellectual responses to all sorts of situations.

The area of socially intelligent agents has for some years acknowledged the importance of autobiographic memories to found the construction of social identity and social interaction in artifacts. For example, Dautenhahn proposed a dynamic systems approach to model autobiographic memories, and the notion of autobiographic agent as the embodied realization of an agent that dynamically reconstructs its individual history over its life-time (Dautenhahn 1996). Nehaniv has used ideas from algebra (semigroup theory) to propose a representation of histories as autobiographies of social agents (Nehaniv 1997). Models of autobiographical memory have also been developed to carry activities such as navigation (Barakova & Lourens 2005), story telling, and role-playing games (Ho & Watson 2006).

The artificial emotion community has also made some attempts at implementing simple versions of notions relevant to the autobiographic self. For example, Velásquez, taking inspiration from (Damasio 1994), implemented emotional memories in a learning pet robot (Velásquez 1998), with the purposes of permitting the learning of secondary emotions as generalizations of primary ones, and of providing markers that influence the robot's decisions. As another example, Ventura and colleagues (2001) have applied Damasio's (1999) concept of the "movie-in-the-brain" to implement a mechanism that allows an agent to establish and learn causal relationships between its actions and the responses obtained from the environment, and to decide courses of action accordingly. The "movie-in-the-brain" mechanism influences decisions on courses of action as follows: the agent stores chunks of sequences of perceptions and actions, together with a measure of their corresponding desirability. When a similar situation is encountered in the future, the agent can make decisions based on its personal experience.

6. Social grounding

In addition to the notions previously mentioned (among many others), stemming from an individual's point of view, robots and other artifacts must incorporate many other mechanisms in order to behave, cognize and emote socially. Let us consider some of the elements that researchers have already started to implement in social artifacts (in particular in robots) which are fundamentally related to social emotions.

6.1 Social motivation

Although different in nature and roles, motivation and emotion are highly intertwined and cannot be considered in isolation from each other. On the one hand, emotions can be seen as “second order” behavior control mechanisms that monitor the motivational system in achieving its goals. On the other hand, emotions modulate motivation (e.g., its intensity) and provide very strong motivation for action.

As in the case of “emotion,” the term “motivation” spans a wide range of phenomena as varied as physiological drives, search for internal (psychological) or external (cultural, social) rewards, or incentives for self-regulation. The mechanisms underlying these different aspects are likely to be very different in fundamental ways, as they have to deal with factors of very diverse nature: biological, psychological, cultural, social, etc. Whereas physiological motivation (drives) are generally explained (at least partially) using a homeostatic regulation approach, models for motivation in self-regulation and social interaction and their connection to emotion are not so well established.

Different models of emotion synthesis for individual artifacts have integrated motivation as one of their elements, taking a homeostatic regulation approach to motivation – see e.g., (Cañamero 1997; Velásquez 1998) for representative seminal work. Inspired from these to a large extent, and partly due to the lack of a better model, architectures that have included motivation for social interactions and emotions have also taken a homeostatic regulation approach to model social motivation (Cañamero & Van de Velde 2000; Breazeal 2002; Blanchard & Cañamero 2006). Motivation for social interactions is thus modeled as a set of drives such as ‘social interaction’, ‘attachment’, ‘fatigue’, ‘stimulation level’, etc. This simplified model can be very productive from a pragmatic point of view. At the architectural level, it has permitted to use the same type of mechanisms to couple motivation and emotion as those used in architectures of non-social agents – a complementary second-order monitoring mechanism, or an element of the homeostatic loop itself, depending on the approach. With respect to social interactions, it has proved capable of producing behavior that engages humans in socially rich emotional interactions and that regulates these interactions (Breazeal 2002).

However, if we want to reach a deeper understanding of the relationships between social motivations and emotions, and a more sound and better grounded coupling of these notions in our artifacts, a finer-grained approach seems to be necessary. The relationships of social motivations and emotions to concepts like competence (in problem-solving, social, etc.), control (external and self-control), self-esteem, coherence and predictability of the environment, self-regulation, and different kinds or rewards, to name a few, need to be explored.

6.2 Theory of mind

The term “theory of mind” is used in developmental psychology to refer to a set of metarepresentational abilities that allow an individual to understand the behavior of others within an intentional framework – or as Tomasello (1999) puts it, to understand others as mental agents. Theory of mind is thus a theory of other people’s minds. It relies on the ability to understand oneself as an intentional agent and to perceive others as being “like me,” to use Tomasello’s expression. A theory of mind thus allows us to correctly attribute “internal states” – perceptions, beliefs, wishes, goals, thoughts, etc. – to others. How would an artifact’s theory of mind affect its social and emotional interactions? Scassellati draws a very eloquent picture:

A robotic system that possessed a theory of mind would allow for social interactions between the robot and humans that have previously not been possible. The robot would be capable of learning from an observer using normal social signals in the same way that human infants learn; no specialized training of the observer would be necessary. The robot would also be capable of expressing its internal state (emotions, desires, goals, etc.) through social interactions without relying upon an artificial vocabulary. Further, a robot that can recognize the goals and desires of others will allow for systems that can more accurately react to the emotional, attentional, and cognitive states of the observer, can learn to anticipate the reactions of the observer, and can modify its own behavior accordingly.

(Scassellati 2000)

We are still very far from achieving this full picture but, recognizing the importance of this notion of social and emotional interactions with robots, various projects have started to implement some basic elements. Kozima, for example, is working on a mechanism for acquisition of intentionality in his humanoid robot *Infanoid* (Kozima 2002), to allow the robot make use of certain methods for obtaining goals. Beginning with “innate” reflexes, *Infanoid* explores a range of advantageous cause-effect associations through its interactions with the environment, and gradually becomes able to use these associations spontaneously as method-goal associations. Scassellati has been working for several years on elements of a theory of mind for a humanoid robot. Taking inspiration from the models proposed by Leslie (1994) and Baron-Cohen (1995), he has been working to specify the perceptual and cognitive abilities that a robot with a theory of mind should employ (Scassellati 2002), focusing initially on the implementation of pre-attentive visual abilities, e.g., to distinguish between animate and inanimate motion, and identify gaze direction for shared attention, and moving to more complex tasks such as mirror self-recognition (Gold & Scassellati 2007).

6.3 Sympathy and empathy

A set of related mechanisms are relevant to the capacity that humans and other social animals have to “connect” with the emotional state of others and to “adopt” it to varying degrees: phenomena like emotional contagion (related to imitation), sympathy, empathy, perspective taking, and prosocial behaviors like helping. The literature has long debated the differences among these phenomena and whether they belong more to the “cognitive” or to the “emotional” realm, in particular in the case of empathy. Recently, a framework has been proposed (Preston & de Waal 2002) that explains all these phenomena within a “perception-action model” applicable to both the cognitive and the emotional domains, and sees empathy as a superordinate category. The cognitive and emotional mechanisms involved in these phenomena vary. Whereas in emotional contagion the observer’s state results “automatically” from the perception of the other’s state, in sympathy – “feeling with” – the observer “feels sorry” for the other, focusing more on his situation than on his physical state; attention is fundamental in empathy – “feeling into” – , where the observer’s state results from the attended perception of the other’s state and, in the framework proposed in (Preston & de Waal 2002), arises from a projection on the part of the observer rather than from a perception of the other’s state. Besides attention, effects of familiarity/similarity, past experience, learning and cue salience are all fundamentally involved in empathy.

Again, researchers in the field of social robots have acknowledged the importance of these phenomena for social understanding and social/emotional interactions (in humans, animals, and artifacts), in particular of empathy as a form of emotional communication that favors the perception of social signals, as discussed for example in (Dautenhahn 1997). Some elements involved in these phenomena are being investigated by different researchers, such as agent architectures to achieve a “sympathetic coupling” (Numaoka 1997), or underlying perceptual mechanisms such as gaze direction for joint attention or cue salience (Scassellati 2002; Breazeal 2002), but we are a long way from artifacts capable of showing empathy. For the moment, all that our emotional and expressive robots can achieve is to “mirror” facial expressions (e.g., Hegel *et al.* 2006) and to produce (something resembling to) “empathic” reactions from humans, as for example reported in (Scheeff *et al.* 2002; Breazeal 2002; Cañamero 2002; Nadel *et al.* 2006).

7. Conclusion

Starting from the idea that the social and the emotional are highly intertwined, this paper has discussed issues around the construction of emotional robots that

have to interact in a social world, and in particular with humans. I have first examined some of the ways in which emotions can enhance social interactions with robots. After considering the debate that opposes “shallow” modeling of observable emotional features versus “deep” modeling that includes an underlying emotional system, I have sketched some ways in which we can anchor emotions in the architecture of robots in order to make emotional interactions meaningful not only to the human, but also to the artifact itself. I have finally discussed some of the cognitive capabilities that robots should incorporate for their emotions to be properly grounded and to give rise to rich social exchanges with humans.

A final comment concerning approaches to design. Most of the projects mentioned take a developmental path to the design and construction of robots with social and emotional capabilities. Ontogenetic development is no doubt necessary for a robot to build up its cognitive and emotional abilities through its interactions with the physical and social environment (Cañamero *et al.* 2006). Again, human development is what we know better (or closer) and it is natural to take it as a model to inspire the design of an artifact that can develop and learn from experience. Designers are facing an enormous challenge here, though, given the complexity of human beings and their developmental process. Starting the robot from scratch is impossible, and the designer has to decide at what level (or from what “primitives”) and according to what theory(-ies) s/he will start to work. Each of the elements discussed in previous sections relies on many other (physical, perceptual and cognitive) capabilities, the implementation of which is equally challenging and not always possible. We thus seem confronted with a dilemma. On the one hand, a developmental approach is necessary both to ground emotions in robots and to gain a better understanding of how emotions develop and interact with other aspects of our cognition and sociality. On the other hand, trying to follow too closely the steps proposed for human development by psychological theories, besides being possible only in a very limited way, can introduce many biases in our model and lead to deadlocks. Although devoid of the richness of these models, complementing this endeavor with an investigation of the development of, and interactions between, emotions, cognition, and sociality using (much) simpler models (more on the bacteria side than on the human side, so to say) could also provide very valuable insights for the understanding of emotions in social species.

Acknowledgement

This chapter revisits, revises and extends my earlier thoughts in Cañamero, L. Building Emotional Artifacts in Social Worlds: Challenges and Perspectives. In

Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium [Technical Report FS-01-02] (pp. 22–30). Menlo Park, CA: AAAI Press.

References

- Aylett, R. & Cañamero, L. (Eds.) (2002). *Proc. AISB'02 Symposium on Animating Expressive Characters for Social Interactions*. The Society for the Study of Artificial Intelligence and Simulation of Behaviour.
- Barakova, E. & Lourens, T. (2005). Spatial Navigation Based on Novelty-Mediated Autobiographical Memory. In J. Mira & J. R. Alvarez (Eds.), *Mechanisms, Symbols, and Models Underlying Cognition. First International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2005* [LNCS 3561] (pp. 356–365). Berlin & Heidelberg: Springer.
- Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: The MIT Press.
- Blanchard, A. & Cañamero, L. (2005). From Imprinting to Adaptation: Building a History of Affective Interactions. In L. Berthouze, F. Kaplan, H. Kozima, H. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov & C. Balkenius (Eds.), *Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, EpiRob 5*, Lund University Cognitive Studies, 123, 23–30.
- Blanchard, A. & Cañamero, L. (2006). Developing Affect-Modulated Behaviors: Stability, Exploration, Exploitation, or Imitation? In F. Kaplan *et al.* (Eds.), *Proc. 6th International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, EpiRob 6*. Lund University Cognitive Studies, 128, 17–24.
- Bower, G. H. (1981). Mood and Memory. *American Psychologist*, 36, 129–148.
- Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: The MIT Press.
- Brooks, R. A. (1991). Intelligence without Representation. *Artificial Intelligence*, 47, 139–159.
- Cañamero, L. D. (1997). Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In W. Lewis Johnson (Ed.), *Proceedings of the First International Conference on Autonomous Agents* (pp. 148–155). New York, NY: ACM Press.
- Cañamero, L. D. (1998). 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] (pp. 49–54). Menlo Park, CA: AAAI Press.
- Cañamero, L. D. (2001a). Building Emotional Artifacts in Social Worlds: Challenges and Perspectives. In L. D. Cañamero (Ed.), *Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium* [Technical Report FS-01-02] (pp. 22–30). Menlo Park, CA: AAAI Press.
- Cañamero, L. D. (Ed.) (2001b). *Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium* [Technical Report FS-01-02]. Menlo Park, CA: AAAI Press.
- Cañamero, L. D. (2001c). Emotions and Adaptation in Autonomous Agents: A Design Perspective. *Cybernetics and Systems: An International Journal*, 32(5), 507–529.
- Cañamero, L. D. (2002). Playing the Emotion Game with Felix: What Can a LEGO Robot Tell Us about Emotion? In K. Dautenhahn, A. Bond, L. Cañamero & B. Edmonds (Eds.),

- Socially Intelligent Agents: Creating Relationships with Computers and Robots* (pp. 69–76). Norwell, MA: Kluwer Academic Publishers.
- Cañamero, L. D. (2003). Designing Emotions for Activity Selection in Autonomous Agents. In R. Trappl, P. Petta & S. Payr (Eds.), *Emotions in Humans and Artifacts* (pp. 115–148). Cambridge, MA: The MIT Press.
- Cañamero, L. (2005). Emotion Understanding from the Perspective of Autonomous Robots Research. *Neural Networks*, 18, 445–455.
- Cañamero, L. & Avila-García, O. (2007). A Bottom-Up Investigation of Emotional Modulation in Competitive Scenarios. In A. Paiva, R. Prada & R. W. Picard (Eds.), *Second International Conference on Affective Computing and Intelligent Interaction, ACII 2007* [LNCS 4738] (pp. 398–409). Berlin & Heidelberg: Springer.
- Cañamero, L. D. & Fredslund, J. (2001). I Show You How I Like You – Can You Read it in My Face? *IEEE Transactions on Systems, Man, and Cybernetics: Part A*, 31(5), 454–459.
- Cañamero, L. & Gaussier, P. (2005). Robots as Tools and Models for Emotion Research. In J. Nadel & D. Muir (Eds.), *Emotional Development* (pp. 235–258). Oxford: Oxford University Press.
- Cañamero, L. D. & Van de Velde, W. (2000). Emotionally Grounded Social Interaction. In K. Dautenhahn (Ed.), *Human Cognition and Social Agent Technology* (pp. 137–162). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Cañamero, L., Blanchard, A., Nadel, J. (2006). Attachment Bonds for Human-Like Robots. *International Journal of Humanoid Robotics*, 3(3), 301–320
- Damasio, A. (1994). *Descartes' Error. Emotion, Reason, and the Human Brain*. New York, NY: Putnam's Sons.
- Damasio, A. (1999). *The Feeling of What Happens. Body and Emotion in the Making of Consciousness*. New York, NY: Harcourt.
- Damasio, A. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. London: Vintage.
- Dautenhahn, K. (1996). Embodied Cognition in Animals and Artifacts. In *Embodied Cognition and Action. Papers from the 1996 AAAI Fall Symposium* [Technical Report FS-96-02] (pp. 27–32). Menlo Park, CA: AAAI Press.
- Dautenhahn, K. (1997). I Could Be You – The Phenomenological Dimension of Social Understanding. *Cybernetics and Systems*, 28(5), 417–453.
- Dautenhahn, K. (1998). The Art of Designing Socially Intelligent Agents: Science, Fiction, and the Human in the Loop. *Applied Artificial Intelligence*, 12(7/8), 573–617.
- Dautenhahn, K., Bond, A., Cañamero, L. & Edmonds, B. (Eds.) (2002). *Socially Intelligent Agents: Creating Relationships with Computers and Robots*. Norwell, MA: Kluwer Academic Publishers.
- Dumouchel, P. (1999). *Emotions: essai sur le corps et le social*. Le Plessis-Robinson, France: Institut Synthélabo/PUF.
- Fellous, J.-M. & Arbib, M. A. (Eds.) (2005). *Who Needs Emotions? The Brain Meets the Robot*. Cambridge, MA: The MIT Press.
- Frijda, N. H. (1995). Emotions in Robots. In H. L. Roitblat & J.-A. Meyer (Eds.), *Comparative Approaches to Cognitive Science* (pp. 502–516). Cambridge, MA: The MIT Press.
- Gold, K. & Scassellati, B. (2007). A Bayesian Robot that Distinguishes “Self” from “Other”. In *Proc. 29th Annual Meeting of the Cognitive Science Society (CogSci2007)*, Nashville, Tennessee, August 1–4, 2007.
- Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42, 335–346.

- Hegel, F., Spexard, T., Wrede, B., Horstmann, G., & Vogt, T. (2006). Playing a Different Imitation Game: Interaction with an Empathic Android Robot. In *Proc. 6th IEEE-RAS International Conference on Humanoid Robots* (pp. 56–61), Genova, Italy, December 4–6, 2006. IEEE Press.
- Ho, W. C. & Watson, S. (2006). Autobiographic Knowledge for Believable Virtual Characters. In J. Gratch, M. Young, R. Aylett, D. Ballin & P. Olivier (Eds.), *Intelligent Virtual Agents. 6th International Conference, IVA 2006* [LNCS 4133] (pp. 383–394). Berlin & Heidelberg: Springer.
- Kozima, H. (2002). Infanoid: A Babybot that Explores the Social Environment. In K. Dautenhahn, A. Bond, L. Cañamero & B. Edmonds (Eds.), *Socially Intelligent Agents: Creating Relationships with Computers and Robots* (pp. 157–164). Norwell, MA: Kluwer Academic Publishers.
- Leslie, A. M. (1994). ToMM, ToBY, and Agency: Core Architecture and Domain Specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 119–148). Cambridge University Press.
- Lewis, M. D & Granic, I. (Eds.) (2000). *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development*. New York, NY: Cambridge University Press.
- Lowe, R. J., Cañamero, L., Nehaniv, C. L. & Polani, D. (2004). The Evolution of Affect-Related Displays, Recognition and Related Strategies. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami & R. A. Watson (Eds.), *Artificial Life IX: Proc. 9th Intl. Conference on Artificial Life* (pp. 176–181). Cambridge, MA: MIT Press.
- Lowe, R. J., Nehaniv, C. L., Polani, D. & Cañamero, L. (2005). The Degree of Potential Damage in Agonistic Contests and its Effects on Social Aggression, Territoriality and Display Evolution. In *Proc. 2005 IEEE Congress on Evolutionary Computation*, Vol. 1 (pp. 351–358), September 2–5, 2005, Edinburgh, UK.
- Nadel, J. & Muir, D. (Eds.) (2005). *Emotional Development: Recent Research Advances*. Oxford, UK: Oxford University Press.
- Nadel, J., Simon, M., Canet, P., Soussignan, R., Blancard, P., Cañamero, L. & Gaussier, P. (2006). Human Response to an Expressive Robot. In F. Kaplan *et al.* (Eds.), *Proc. 6th International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Epi-Rob 6*, Lund University Cognitive Studies, 128, 79–86.
- Nehaniv, C. L. (1997). What's Your Story? – Irreversibility, Algebra, Autobiographic Agents. In *Socially Intelligent Agents. Papers from the 1997 AAAI Fall Symposium* [Technical Report FS-97-02] (pp. 150–153). Menlo Park, CA: AAAI Press.
- Numaoka, C. (1997). Innate Sociability: Sympathetic Coupling. In K. Dautenhahn (Ed.), *Socially Intelligent Agents. Papers from the 1997 AAAI Fall Symposium* [Technical Report FS-97-02] (pp. 98–102). Menlo Park, CA: AAAI Press.
- Ortony, A. (2003). On Making Believable Emotional Agents Believable. In R. Trappl, P. Petta & S. Payr (Eds.), *Emotions in Humans and Artifacts* (pp. 193–211). Cambridge, MA: The MIT Press.
- Ortony, A., Clore, G. L. & Collins, A. (1988). *The Cognitive Structure of Emotions*. New York, NY: Cambridge University Press.
- Paiva, A. (2001). *Affective Interactions: Towards a New Generation of Computer Interfaces* [LNCS/LNAI 1814] Berlin & Heidelberg: Springer-Verlag.
- Paiva, A., Prada, R. & Picard, R. (Eds.) (2007). *Affective Computing and Intelligent Interaction, Second International Conference, ACII 2007* [LNCS 4738] Berlin & Heidelberg: Springer.

- Pelachaud, C. & Cañamero, L. (Eds.) (2006). Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids. Special issue of the *International Journal of Humanoid Robotics*, 3(3).
- Prendinger, H. & Ishizuka, M. (2004). *Life-Like Characters: Tools, Affective Functions, and Applications*. Berlin: Springer.
- Preston, S. D. & de Waal, F. B. M. (2002). Empathy: Its Ultimate and Proximate Bases. *Behavioral and Brain Sciences*, 25(1), 1–20.
- Reeves, B. & Nass, C. (1996). *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY: Cambridge University Press/CSLI Publications.
- Sacks, O. (1986). *The Man Who Mistook His Wife for a Hat*. London, UK: Picador.
- Scassellati, B. (2002). Theory of Mind for a Humanoid Robot, *Autonomous Robots*, 12, 13–24.
- Scheeff, M., Pinto, J., Rahardja, K., Snibbe, S. & Tow, R. (2002). Experiences with Sparky, a Social Robot. In K. Dautenhahn, A. Bond, L. Cañamero & B. Edmonds (Eds.), *Socially Intelligent Agents: Creating Relationships with Computers and Robots* (pp. 173–180). Norwell, MA: Kluwer Academic Publishers.
- Tao, J., Tan, T. & Picard, R. (Eds.) (2005). *Affective Computing and Intelligent Interaction, First International Conference, ACII 2005* [LNCS 3784]. Berlin & Heidelberg: Springer.
- Tomasello, M. (1999). *The Cultural Origins of Social Cognition*. Cambridge, MA: Harvard University Press.
- Trapp, R. & Petta, P. (Eds.) (1997). *Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents* [LNCS/LNAI, Vol. 1195]. Berlin & Heidelberg: Springer-Verlag.
- Velásquez, J. D. (1998). Modeling Emotion-Based Decision-Making. In *Emotional and Intelligent: The Tangled Knot of Cognition. Papers from the 1998 AAAI Fall Symposium* [Technical Report FS-98-03] (pp. 164–169). Menlo Park, CA: AAAI Press.
- Ventura, R., Custódio, L. & Pinto-Ferreira, C. (2001). Learning Courses of Action Using the “Movie-in-the-Brain” Paradigm. In L. Cañamero (Ed.), *Emotional and Intelligent II: The Tangled Knot of Social Cognition. Papers from the 2001 AAAI Fall Symposium* [Technical Report FS-01-02] (pp. 147–152). Menlo Park, CA: AAAI Press.
- Wehrle, T. (2001). The Grounding Problem of Modeling Emotions in Adaptive Systems. *Cybernetics and Systems*, 32(5), 561–580.
- Wehrle, T. & Scherer, K. (1995). Potential Pitfalls in Computational Modeling of Appraisal Processes: A Reply to Chwelos and Oatley. *Cognition and Emotion*, 9, 599–616.

Dynamic models of multiple emotion activation

Valeria Carofiglio, Fiorella de Rosi and Roberto Grassano

1. Introduction

The ability to show affective behaviour is recognized as one of the essential ingredients of believability in Embodied Animated Characters. If artificial agents cannot be built to ‘feel’ emotions (at least as far as internal bodily changes are concerned), they must at least be able to simulate this condition in their external appearance. Shallow or inner aspects of behaviour may be influenced by emotions, such as facial expressions, gestures, movement but also decision making, argumentation style, or instructional strategies (Sillince & Minors 1991; Forgas 2000; Gmytrasiewicz & Lisetti 2000; Staller & Petta 2001).

Many of the recent studies aimed at introducing some form of affective behaviour in Human-Computer Interaction originated from the seminal works of Carbonell (1980), Oatley and Johnson-Laird (1987), Ortony, Clore and Collins (1988) and others. In particular, Ortony and colleagues suggested a categorization of emotions that became the starting point of the large majority of projects. Emotion activation modelling methods were also inspired by the emotion activation rules suggested by Ortony (1988) and by the identification of variables influencing this process proposed by Elliott and Siegle (1993).

The way complex phenomena like emotions may be modelled depends on the envisaged application area. If, for instance, the application concerns a 2D embodied character that is sketchy in its appearance and is expected to show a limited range of expressions, a refined modelling approach is probably not needed. This is the case with Microsoft-Agents, which are not able to show single or multiple emotions with graded intensities.

In other cases, more refined representations are both needed and possible. Typically, 3D characters are designed to be highly realistic; their face and bodies may be manipulated so as to show the large variety of expressions that are displayed by humans. A requirement for high realism also occurs when the goal of realism and believability concerns the social behaviour of characters rather than (or

in addition to) their external appearance: for instance, when they are expected to undertake ‘natural conversations’ with the user. In the latter case, some knowledge of the reasons why an emotional state was activated (in the agent and in the user) is useful to achieve consistency and naturalness in the behaviour, both at a given time instant and over time. Models have then to be refined enough to represent the cognitive and social aspects of emotion activation and decay. For instance:

- which beliefs and goals contribute to activating each emotion,
- how differently agents react to similar situations according to their personality and to the context in which the event occurs,
- which variables affect emotion intensities and, finally,
- how emotions may combine and vary over time.

In these cases, a simple representation of the correspondence between event and emotion is not sufficient and a finer granularity of knowledge is needed. In addition, the evolution over time of the affective state has to be represented.

In this paper, we discuss this problem in general terms, starting from Picard’s ‘Marathon example’ (Picard 1997, p. 171). We then propose a method to represent cognitive models of emotion activation and decay which deals with uncertainty and goal values.

2. Multiple emotions

In the section about ‘Pure vs. Mixed’ emotions of her book on affective computing, Rosalind Picard introduces the following example: After winning a Marathon, Uta, a professional runner ‘described feeling tremendously *happy* for winning the race, *surprised* because she believed she would not win, somewhat *sad* that the race was over and a bit *fearful* because during the race she had acute abdominal pain’ (Picard 1997, p. 171). A runner’s friend present at that Marathon but unable to participate himself was probably *happy* for her because she won the race, although a little *envious* for not being able to participate in it and *sorry* for seeing her so tired at the end.

How is it that these two persons reported feeling this different mixture of emotions? Clearly, the main source of difference is due to the different structure of beliefs and goals in their minds. In the runner, the *intensity* of *fear* during the race was probably related, at the same time, to the *importance* she assigned to her goal of winning it and to variations in the *probability* of achieving this goal, which she dynamically revised during the race. The importance of this goal also affected the intensity of *happiness* (or *satisfaction*) of achieving it, while *surprise* was probably due to a difference between the likelihood she attached to achieving the goal

at the beginning of the race and the final result. The *sadness* that the event was over might be a mixed emotion in its turn, some combination of *nostalgia* for a pleasant past event and *hope* to be again in a similar situation, in the future. The mixing of emotions in the runner's friend was probably due to a mixing of goals of approximately equal weight. *Happy-for* was due to his desire to achieve 'the good of his friend' (attached to her winning the race); *sorry-for* was due to his desire to 'preserve her from harm' (illness, in this case); *envy* was due to his desire to 'dominate her', in a way.

It therefore seems that differences between the two persons in the example are due to differences in their beliefs, the goals they want to achieve, the weights they assign to achieving them and the structure of links between beliefs and goals. Variations of these measures over time seem to govern cognitively-generated emotions. Picard evokes the *generative mechanism* as the key factor for distinguishing between emotions that may *coexist* (by mixing according to a 'tub of water' metaphor) and emotions that *switch* from each other over time (by mixing according to the 'microwave oven' metaphor). She suggests that co-existence may be due, first of all, to differences in these generative mechanisms. But they may be due, as well, to differences in the decay speed between emotions that were generated by the same mechanism at two distinct time instants: for instance, 'primary' emotions, like *fear*, and cognitively-generated ones, like *anticipation*.

To represent the two ways in which emotions may mix, the modelling formalism adopted should therefore be able to represent their generative mechanism, the intensity with which they are triggered and the way this intensity decays over time. We claim that *dynamic belief networks* (DBNs) are an appropriate formalism to achieve these goals. We show that this formalism is able to represent the dynamic arousal, evolution and disappearance of emotions and the way these phenomena are influenced by personality factors and social context. We illustrated in detail the logic behind this modelling method in another paper, in which we discussed the advantages it offers in driving the affective behaviour of an Embodied Conversational Agent that we are building in the scope of the European project Magicster¹ (de Rosis *et al.* 2003).

In this contribution, we summarise the main features of our modelling method, to focus our discussion on how the two metaphors of mixing emotions proposed by Picard may be represented. We restrict our analysis to event-based emotions, in the OCC classification (Ortony *et al.* 1988). We briefly describe, in

1. Magicster is a European Research Project funded in the scope of the Information Societies Technology Programme IST-1999-29078. It included the following partners: ICCS, University of Edinburgh, UK; DFKI, Germany; SICS, Sweden; AvatarMe, UK; DIS, University of Roma 'La Sapienza', Italy; Dipartimento di Informatica, University of Bari, Italy.

Section 3, what DBNs are and how they may be employed as a monitoring tool to represent emotion triggering. We then focus our analysis on how differences in the cognitive generative mechanism of emotions may be represented with DBNs, with the aim of establishing a correspondence between these differences and the two mixing metaphors proposed by Picard. We finally contrast our method with some of the alternative approaches that have been proposed recently.

3. Emotion triggering with DBNs

As we anticipated, our departure point is that emotions are activated by the belief that a particular important goal may be achieved or threatened. So, our simulation is focused on the *change* in the belief about the achievement of (or threat to) goals of an agent A over time. In our monitoring system, the cognitive state of A is modelled at the time instants $\{T_1, T_2, \dots, T_i, \dots\}$. Events occurring in the time interval (T_i, T_{i+1}) are observed to construct a probabilistic model of the agent state at time T_{i+1} and reason about emotions that might be triggered by these events. DBNs are a well known formalism for representing dynamic phenomena in conditions of uncertainty. They are based on the idea that time is divided into time slices, each representing the state of the modelled world at a particular instant T_i . This state is described by means of a static belief network: $\text{World-}T_i$, with its observable state variables $\text{State-Obs-}T_i$. When DBNs are employed for monitoring purposes, two consecutive time slices are linked by arrows between the domain variables that have to be monitored. When something changes in the world, an event $\text{Event-}T_i\text{-}T_{i+1}$ occurs, that is observed through the variables in $\text{Change-Obs-}T_i\text{-}T_{i+1}$. The network is then extended for an additional time slice T_{i+1} . As a consequence, its structure and the probabilities of its nodes usually change (Pearl 1988). To avoid explosion in the complexity of the network (and therefore in the uncertainty propagation algorithm), pruning of time slices and of network parts is performed after a new observation is added to the model, with a mechanism of roll-up (Nicholson & Brady 1994).

Figure 1 shows the general structure of our model of emotion activation, which includes the following static components:

- $M(T_i)$ represents the agent's mind at time T_i , with its beliefs about the world and its goals;
- $\text{Ev}(T_i, T_{i+1})$ represents an event occurred in the time interval (T_i, T_{i+1}) , with its causes and consequences;
- $M(T_{i+1})$ represents the agent's mind at time T_{i+1} ;
- $\text{Em-feel}(T_{i+1})$ represents activation of a particular emotion in the agent at time T_{i+1} .

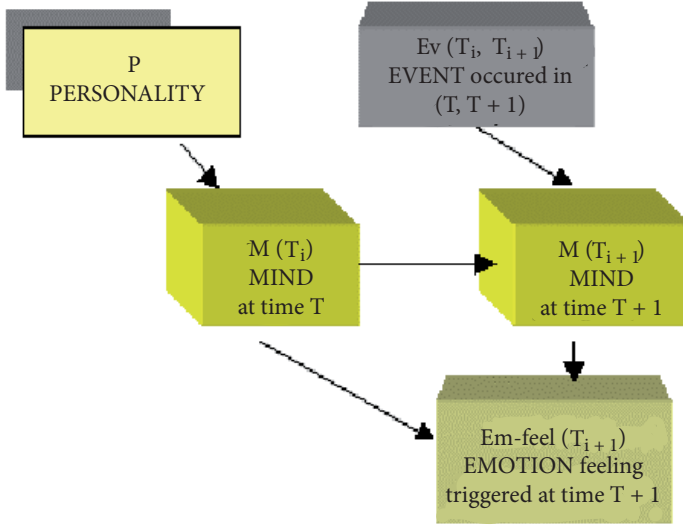


Figure 1. Outline of our emotion monitoring system

$M(T_{i+1})$ depends on $M(T_i)$ and the event occurring in the interval (T_i, T_{i+1}) . The feeling of emotion depends on both $M(T_i)$ and $M(T_{i+1})$. We calculate the intensity of emotions as a function of two parameters: (1) the *uncertainty* in the agent's beliefs about the world and, in particular, about the possibility that some important goal is achieved or threatened, and (2) the *utility* assigned to this goal. In more depth, if:

- A denotes the agent; G_i a high-level goal and $Ach-G_i$ the achievement of this goal;
- $Bel A Ach-G_i$ denotes A 's belief that the goal G_i will be achieved;
- $P(Bel A Ach-G_i)$ and $P^*(Bel A Ach-G_i)$ denote, respectively, the probabilities that A attaches to this belief, before and after the event Ev occurred;
- $W_A(Ach-G_i)$ denotes the weight that A attaches to achieving G_i , then the *variation of intensity in the emotion* (ΔIe) may be computed by applying the utility theory (Pearl 1988), as follows:

$$\Delta Ie = [P^*(Bel A Ach-G_i) - P(Bel A Ach-G_i)] * W_A(Ach-G_i)$$

In other words, ΔIe is the product of the change in the probability that G_i will be achieved, times the weight of this goal. For negatively-valenced emotions (such as fear, sorry for etc) we represent, instead, the probability that a goal G_i will be threatened ($Thr-G_i$). In the following section we will discuss the relations that may hold between achievement of and threats to some goals.

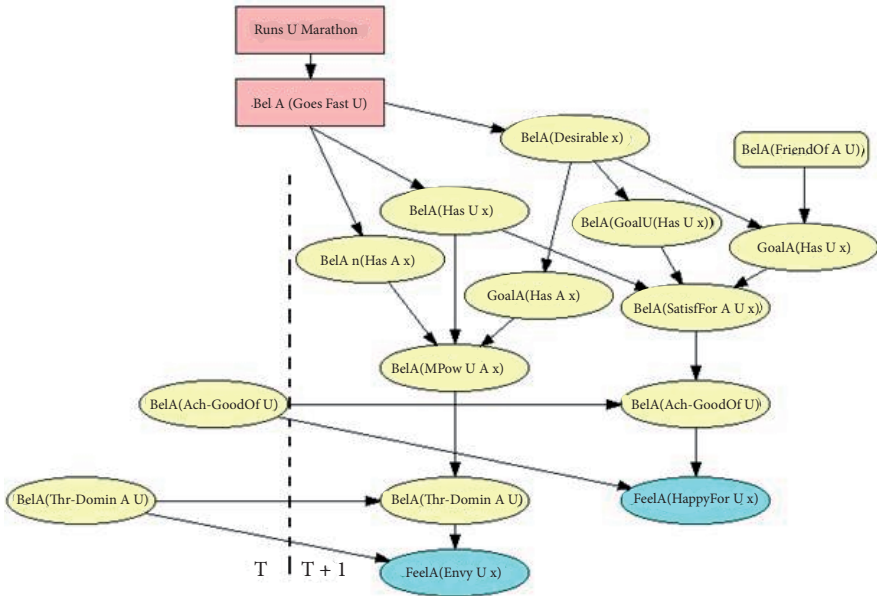


Figure 2. The DBN for envy and happy-for

4. Event-based emotion categories

In Ortony, Clore and Collins' categorization (Ortony *et al.* 1988), emotions may be activated by events, actions and aspects of objects. Event-based emotions are elicited from events occurring in the social environment in which the agent is situated and may be grouped in four main subcategories: Fortune-of-others, Prospect-based, Well-being and Confirmation.

Fortune-of-others emotions (sorry-for, happy-for, envy and gloating) may be represented as points in the two-dimensional space ('desirability of the event', 'empathic attitude'). Happy-for and envy apply to desirable events while sorry-for and gloating apply to undesirable ones; happy-for and sorry-for are driven by an empathic attitude, while gloating and envy are driven by a non empathic (or even contrasting) one.

Figure 2 shows the dynamic belief network that models how happy-for and envy may be activated in an agent A that assists to her friend U running the Marathon. This model shows that happy-for is triggered after believing that U will win the race or at least will complete it with a good time. If this event occurs, the probability of the belief that the goal of getting the good of others will be achieved (Ach-GoodOf U) increases. The intensity of this emotion depends on how this probability varies when evidence about the mentioned desirable event is

propagated in the network. It depends, as well, on the weight the agent attaches to achieving that goal; this weight is, in its turn, a function of the agent's personality: it is high for altruistic people, low for egoistic ones. The emotion of sorry-for is triggered by the same variables as happy-for. However, some of these variables have an opposite sign than in the case of happy-for because they are affected by undesirable events: in our example, when the runner seems to be suffering. The high-level goal involved in triggering sorry-for is preserving others from harm (Thr-PresBad U); this goal is negatively correlated to getting the good of others so that happy-for and sorry-for cannot be triggered by the same event. Gloating is related to envy by a similar triggering mechanism: the high-level goals involved are, in this case, desiring the harm of others and preserving self from harm.

Which fortune-of-others emotions may mix and how? In our cognitive model of emotion activation, a correspondence may be established between cognitive generation of emotions and the set of beliefs and goals that influence the probability of achieving (or threatening) the goals that govern their activation. Let us consider again the BN for activation of envy and happy-for that is shown in Figure 2. As we said, the goal involved in the activation of happy-for is getting the good of others (in particular, of U): happy-for is triggered by the belief that achieving this goal (Bel A (Ach-GoodOf U)) increases over a given threshold. The root nodes of this sub network, that may influence variation of this probability and are directly influenced by the considered event, are the following:

Bel A (Has U x): the agent A believes that U has something x

Bel A (Desirable x): x is desirable.

If envy is considered instead, the threatened goal is dominating others and the root nodes are the same as for happy-for, plus the additional belief:

Bel A not (Has A x): the agent A has not x.

The cognitive generation mechanism is therefore directly represented in the root nodes of the subnets affecting goal achievement or threat, in the Agent's mental state. Two emotions can coexist if and only if their cognitive generation mechanisms are compatible: that is, if all the root nodes in their activation subnets take compatible values. According to this model, happy-for and envy are examples of potentially coexisting emotions. As we saw in Figure 2, their generation subnets share a set of root nodes which take compatible values when evidence about some observed event is propagated; in particular, the nodes Bel A (Desirable x) and Bel A (Has A x) are both true. Therefore, agents who are moderately altruistic and moderately dominant may be moderately envious and moderately happy-for at the same time, when they come to know that a desirable event occurred to a

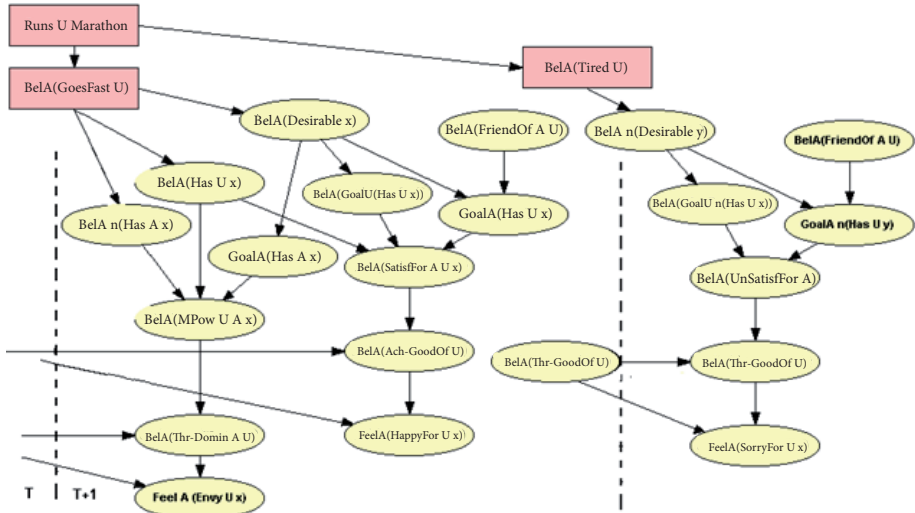


Figure 3. The DBN for happy-for and sorry-for

friend. For similar reasons, sorry-for and gloating may coexist, under particular circumstances (we do not show the DNB for this case).

The same mixing metaphor does not apply to other emotion combinations within the category ‘fortune-of-others’ because they are triggered by a different value of desirability of the occurring event and, therefore, by incompatible values of some of the root nodes they share in their generation subnets. However, it might happen that the same event activates, in two different time instants, different kinds of beliefs. Persons may see positive and negative consequences of the same event at different times. In our example (see Figure 3), the agent may notice that ‘U is going fast’ in the interval (T_i, T_{i+1}) , hence feeling happy-for – and, as we said, maybe also a bit of envy. A may subsequently notice (in T_{i+1}, T_{i+2}) that ‘U seems to be tired’: if the second event occurs before the effect of the previous one has disappeared, a sorry-for will co-occur with the previous emotion, according to a ‘tub of water’ effect. Otherwise, the Agent will switch between the two (positive and negative) emotional states, according to a ‘microwave oven’ effect.

More or less fast switching in time may occur in other emotion categories, from *Prospect-based* emotions (fear, hope) to *Well-being* (distress, joy) or *Confirmation* emotions (disappointment, relief). Beliefs about achievement of or threats to high-level goals (‘Getting the Good of Self’ or ‘Preserving Self from Bad’) are involved in these cases. Switching from one emotion to another is due to a change in the probability of the belief that a (desirable or undesirable) event will occur, is occurring or occurred. This change may be due to observation of different pieces of evidence originating from this event at different times. In the Marathon ex-

ample, switching from fear to joy to relief is closely related to the probability with which the runner believes she will win the race; this probability changes from time to time, according to the progress of the race. A ‘microwave oven’ metaphor may therefore be applied to the mixing of emotions originating from a change of value of some belief at different time instants, in the three categories. The same happens for emotions within the Prospect-based, the Well-being or the Confirmation category, which cannot coexist but between which the Agent may switch. For instance, the Marathon runner was, in different time instants, hopeful of winning the race and fearful about its possible consequences on her health. According to similar considerations, one may conjecture about the mixing of emotions belonging to different categories: for instance, happy and happy-for may coexist, as well as sorry-for and relief, and so on.

5. A tool for simulating mixed emotions

A good way to validate representativeness of a model is to check its behaviour in some ‘typical’ application domain. To test our emotion activation modelling method, we built a test bed with a graphical interface (that we called Mind-Tested) that we applied to domains of various complexities. Mind-Tested enables us to build, test and refine a model through an iterative process that comprises a building step and a testing step. In the *building step*, the following components are set up:

1. The *agent’s mental state*, as a belief network.
2. For every application domain, the *events that may occur* in this domain, again as belief networks.
3. The set of *personalities* the agent may take: every personality is a psychologically plausible combination of traits; every trait corresponds to assigning a weight (in the 0–10 scale) to a goal-node in the agent’s mind. Figure 4 shows an example of personality: an ‘other-centred’ agent gives a high weight to the goals of gaining the good of others and preserving them from harm. If the agent is also optimistic, it gives a high weight also to the goal of achieving the future good of self.
4. The set of *contexts* in which interaction with the environment may take place; a context corresponds to assigning a value to one or more context-nodes in the Agent’s mind. Figure 5 shows an example in which the ‘social’ context is set up, to be employed as shown (for instance) in Figures 2 and 3 for the models of envy, happy-for and sorry-for. A friendship context involves adopting the interlocutor’s goal as far as possible.

5. A model of the relationship between emotions and goals in the agent's mind: in these models, the way every emotion decays is mirrored in the conditional probability table associated with the link between goal nodes at times T_i and T_{i+1} . The higher the probabilities assigned to the (true, true) and the (false, false) combinations, the slower the emotion decay (as shown in Figure 6). Decay time may be defined differently for the different emotions and may be varied according to another trait of the agent's personality: how persistent it is in its mood.

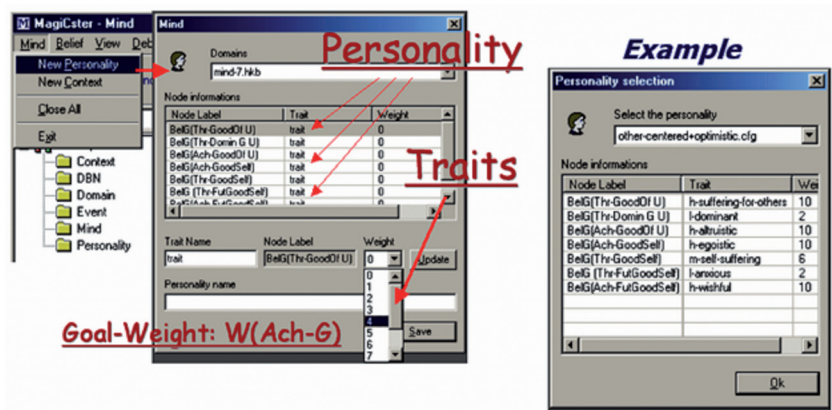


Figure 4. Defining personality traits

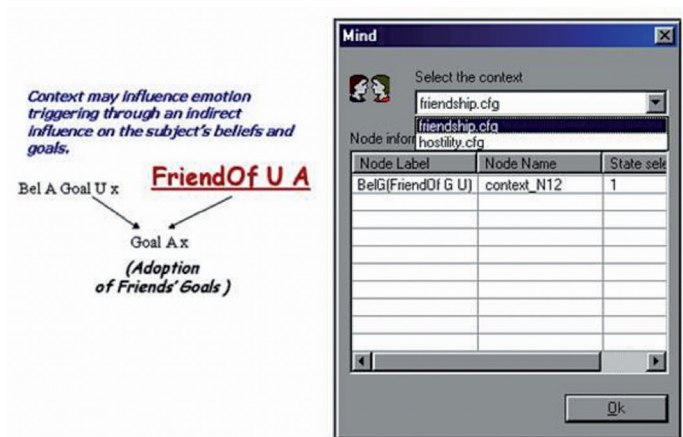


Figure 5. Defining contexts

In the *testing step*, a graphical interface enables selecting a personality for the Agent, a context in which interaction occurs, a domain and a threshold for emotion activation (Figure 7). A sequence of events is then fired and the way the agent's emotional state changes with time as a consequence of these events is monitored (in tabular and in graphical form) in the bottom frame.

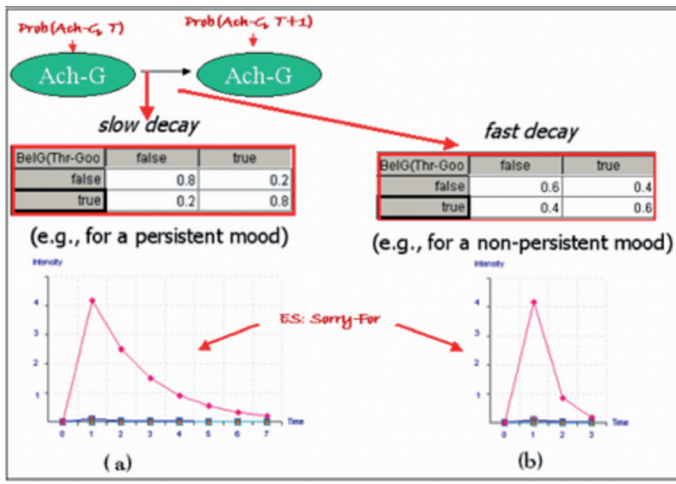


Figure 6. Setting emotion decay

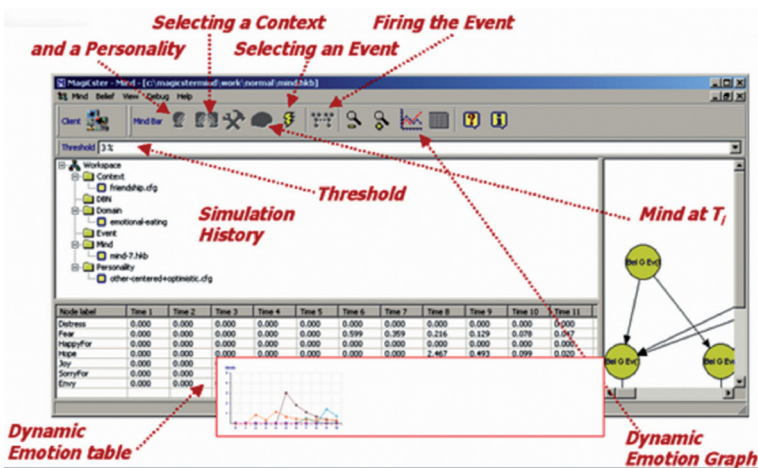


Figure 7. Interface in the testing phase

6. An Example

Let us consider again Picard's marathon example, to imagine the following dialog between the agent (A) and his friend Uta (U):

- U0: I'm OK. I'm planning to run the Boston Marathon.
A1: Yes, I know. Are you in good training shape?
U1: My coach says so.
A2: Good, I'm happy for you.
U2: And what about you? Are you coming to Boston too?
A3: No, unfortunately I can't. Do you feel OK?
U3: I'm not at my best. I got injured on my last run.
A4: Oh! I'm sorry for you!
U4: Nothing serious.
A5: And now?
U5: Now I'm fairly OK. What about a small run together?
A6: Sure, with pleasure!

The agent talks at times $T_0, T_1, T_2, \dots$. The interlocutor's moves, in the intervals between two time instants, are the events that may trigger an (exogenous) emotional state in the agent.² In the example (see Figure 8), the agent feels a mixture of happy-for and envy at time T_2 (move A2), after Uta declares that she will run the next Boston Marathon (a desirable event). At time T_4 (move A4), after Uta declares not to be in her best conditions, the agent feels sorry-for, which may overlap to some still lasting happy-for and envy. Finally, at time T_6 (move A6), joy is triggered from the desirable event (to A) of going to run with Uta; this emotion overlaps with a very low (due to the decay) happy-for and sorry-for. Once felt, an emotion may be displayed or hidden, according to the context in which interaction occurs (De Carolis *et al.* 2001). Emotions are displayed in the content and the style of the agent moves. At move A2: '*Good, I'm happy for you!*' the agent manifests its happy-for while it hides its light envy. At move A4: '*Oh, I'm sorry for you!*' it manifests its sorry-for. At move A6, '*With pleasure!*' it manifests its joy. Although multiple emotions are felt, only one of them is displayed at every move,

2. The description of how we simulate the dialog is out of the scope of this contribution and may be found in (Cavalluzzi *et al.* 2003). In this system the user moves are translated into symbolic communicative acts by a simple keyword analysis. They are then recognized as 'leaf-nodes' in an Event-BN and are propagated in the DBN to assess their emotional impact on the agent. Agent moves are enriched with affective tags according to the results of the emotion activation component and may be expressed as natural language strings or may be given as an input XML file to an embodied character.

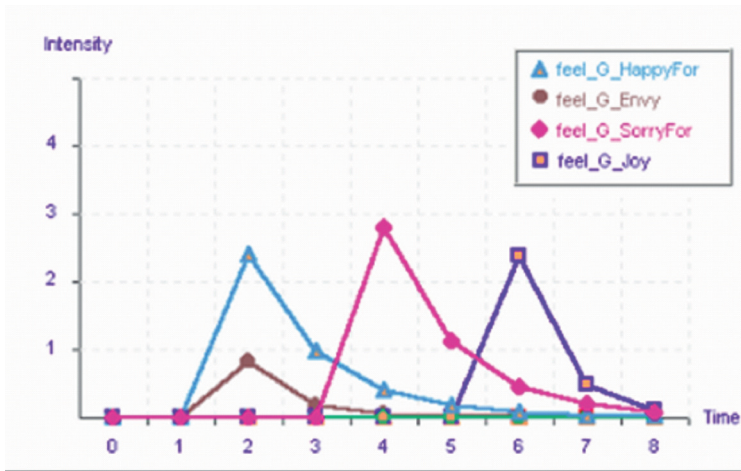


Figure 8. Emotion activation in the Marathon dialogue

in this example. However, more complex sentences might enable displaying more complex emotional states (as, for instance, in Walker *et al.* 1997).

When the agent takes the appearance of an embodied animated character, it may show its emotional state through facial expression, gesturing, head and body movements etc, as we describe in (de Rosis *et al.* 2003). In this case, multiple emotions reflect into composite body expressions (Kaiser & Wehrle, this volume; Coulson, this volume). The same model may be applied to recognize rather than generate emotional states. When the interlocutor manifests some emotion, the agent may try to guess the causes of this state: it may try to infer the set of beliefs and goals that probably produced it, through some form of ‘emotional plan recognition’. To do this, it employs knowledge about the context, the interlocutor’s personality and the events occurred in previous steps of the dialog.

7. Related work

The work discussed here is part of a flourishing of proposals on how to model emotion elicitation, expression and recognition. They build upon a common core of psychological theories (Ortony *et al.* 1988; Elliott & Siegle 1993; Oatley & Johnson-Laird 1987, to cite only the main ones) and differ in the methods employed to formalise them. Two main modelling methods have been applied so far.

The first modelling approach follows the line traced by Ortony and his colleagues in the late 1980s (Ortony *et al.* 1988), by proposing mathematical functions to combine a large number of (numerical) parameters in a measure of emotion

intensity. In the Affective Reasoner, Elliott lists the variables that may influence the intensity of an agent's affective state (Elliott & Siegle 1993) without proposing any specific 'emotion-intensity calculation function'. The computational model of the Affective Reasoner was later developed to formalise emotional reasoning in Emile (Gratch & Marsella 2001). The main idea behind this system is to link emotion elicitation to appraisal of the state of plans in memory rather than directly to events. Emotion intensities are measured as a function of the goal importance and of the probability of goal attainment; this is evaluated, in its turn, by an analysis of the plans that could bring that goal about. In (Prendinger *et al.* 2002) this function is a logarithmic combination of the sum of exponentials of a number of variables. The intensity of joy depends on the desirability of the triggering event; the intensity of happy-for is a function of the interlocutor's happiness and of the 'degree of positive attitude' of the Agent towards the interlocutor, and so on. If the number of variables is low, emotion intensity depends on the 'strongest' variable. Personality influences the filtering process, in which it is (consciously or unconsciously) decided whether the emotion should be expressed or suppressed. The main limitation of this method is, in our view, in the combination of several heterogeneous measures in a single function: when different scales are employed, the effect of their combination cannot be foreseen precisely. In addition, the range of variability of the function may go outside the range of the variable that this function should measure, with the need to normalise the obtained value or to cut off values out of this range.

The second approach to affective modelling is based on representing the cognitive aspects of emotion triggering. Belief Networks (BNs) seems to be an ideal formalism in this case. They have a long tradition as a user modelling method in conditions of uncertainty (Zukerman & Albrecht 2001) and have been applied also to model generation and recognition of emotions from the expression of the Agent (Ball & Breese 2000). Dynamic BNs (DBNs) have the additional advantage of enabling representation of the dynamic aspects of this phenomenon.

Our modelling method follows this second line of research. The main difference from the cited work is in the granularity of the knowledge represented. Rather than representing directly the effect of events on the agent's mind, we build a fine-grained cognitive structure in which the appraisal of what happens in the environment is represented in terms of its effects on the agent's system of beliefs and goals. We only employ two measures (uncertainty of beliefs and utility of goals), while Elliott and Siegle propose a different measure for each variable. However, there is a close relationship between the aspects of the phenomenon that are represented in the two cases. Elliott and Siegel's 'importance of achieving a goal/not having a goal blocked' corresponds to the 'weights' we attach to goal achievement or threat. All variables associated with our 'belief' nodes correspond

to their ‘simulation event’ variables (for instance, how appealing a situation is) and ‘stable relationship variables’ (for instance, how desirable an event is, how friendly the agent is with her interlocutor). Rather than attaching to these variables an integer value according to some established scale (as Elliott and Siegle propose), we discretise them in a limited number of values with a probability distribution. By attaching conditional probability distributions to the network’s links, we give these links a strength value in probabilistic terms.

After the pioneering study by Ball and Breese, Bayesian Networks became the most popular formalism for dealing with uncertainty in the interpretation of various ‘signs’ of emotions, e.g. gesturing (Abbasi 2007), facial expressions (Datu & Rothkrantz 2004), and multiple modalities (Sebe *et al.* 2005; Li & Ji 2005).

Due to the awareness of researchers of the importance of uncertainty in emotion activation, Bayesian Networks have also played a key role in cognitive emotion modelling.³ Conati (2002) proposed a model based on Dynamic Decision Networks to represent the emotional state induced by educational games and how these states are displayed. The emotions potentially represented in this model (reproach, shame and joy) belong to the OCC classification. Some personality traits are assumed to affect the student’s goals. The grain-size of representation is not very fine, as among the various attitudes that may influence emotion activation (first and second order beliefs and goals, values etc) only goals are considered in the model. Subsequently, the same group (Conati & McLaren 2005) described how they refined their model by adding new emotions and learning parameters from data collected from real users. Hudlicka (2003) modeled in MAMID the effects of multiple, interacting emotions and personality traits on the cognitive processes mediating decision-making. Prendinger and Ishizuka (2005) describe an artificial agent that empathizes with a relaxed, joyful or frustrated user by interpreting physiological signals with the aid of a probabilistic decision network; these networks include representation of events, agent’s choices and utility functions. Ji and colleagues (2004) combined a Dynamic Bayesian Network that recognizes affective states from sensory data with an affective cognitive model that traces and predicts the impact of affect on the cognitive process. Hernandez and colleagues (2005) endowed an intelligent tutoring system with an affective behaviour model based on a decision network that integrates the student’s cognitive and affective state, and the tutorial situation, to decide the best pedagogical action. In the model proposed by Guo Jiang and colleagues (2007), activation of emotions was considered as a Markov process describing the relation of emotions, external incentives and personality.

3. We mention here only some of the work developed after our presentation at the AISB 2002 Symposium, of which this chapter is an elaboration.

8. Discussion and conclusion

The advantages of our formalism compared to the related work mentioned in the previous section lie in the possibility of representing the Agent's affective state in terms of multiple emotions and how this state changes over time, as a consequence of 'endogenous' or of 'exogenous' events. The role of two categories of factors in influencing emotion elicitation and decay may be considered:

1. Temperament and personality: how they affect the Agent's propensity to feel and show emotions, by influencing the weights it assigns to achieving its goals, the threshold for emotion activation and the resistance to change its emotional state in absence of new specific stimuli;
2. Social context: how the Agent's relationship with the context influences emotion triggering by influencing its beliefs and goals.

Its main limitations arise from the difficulty of calibrating parameters in model building. Ideally, parameters in BNs should be introduced after some empirical analysis of situations. This is not done frequently in applications of BNs to user modelling, due to the difficulty of collecting data about 'hidden' aspects of the users behaviour (their beliefs, goals, preferences etc) and their relationships with the 'observable' ones (actions performed). The alternative is to employ some variant of the 'expert system approach', by asking experts to estimate parameters subjectively. In this case, calibration and sensitivity analysis are performed by testing the models in various situations (agent personalities and social contexts) and by validating (again, subjectively) the results produced by the model. A testbed is of great help in this work. We refined the structure of our DBNs and their parameters through a systematic evaluation of the kinds of dialogs we want our artificial agent to make and through an evaluation of the believability of these dialogs.

By comparing our model with the diverse BN-based emotion models developed since 2002, we still believe that it presents many valuable strong points, especially in the context of affective dialogues. Among other features, we can mention its ability to model the role of personality and social relationship in emotion activation and time decay, its ability to represent the various ways in which emotions may mix up with different intensities and, finally, its potential of being used, as well, for interpreting the reason why an emotion is displayed (for a detailed treatment of this topic see Carofiglio & de Rosis 2005).

We plan to extend our models to other groups of emotions in the OCC classification. One might suspect that a modelling method that is based on goal monitoring is not suited to model action-based or norm-based emotions. We are reasonably confident that it is not so. Take, for instance, the example of shame, which

is an action-based emotion in the OCC categorization. According to Castelfranchi (2000), activation of shame is related to the belief that the goal of showing a good social image might be threatened. A similar activation mechanism may be hypothesized for guilt, in which the goal of showing a good self-image is involved (Miceli 1992). Of course, beliefs about threats to these goals are conditioned through the evaluation of some standard characteristics or behaviour. However, this is true also for event-based emotions, in which the subjective evaluation of 'desirability' of events affects the intensity of emotions activated in a given circumstance. Coupling goal monitoring with evaluation of events, actions and object attributes, therefore provides us the information we need to estimate emotion intensity and trend in our agents.

References

- Abbasi, A. R. (2007). A Bayesian network approach to interpret affective states from human gestures. In R. Cowie & F. de Rosi (Eds.), *The Second International Conference on Affective Computing and Intelligent Interaction (ACII 2007), Proceedings of the Doctoral Consortium*. Lisbon, Portugal, September 12–14, 2007. <http://www.di.uniba.it/intint/DC-ACII07/ListOfPapers.html>.
- Ball, G. & Breese, J. (2000). Emotion and personality in a conversational agent. In S. Prevost, J. Cassell, J. Sullivan & E. Churchill (Eds.), *Embodied Conversational Agents* (pp. 89–219). Cambridge, MA: The MIT Press.
- Carbonell, J. C. (1980). Towards a process model of human personality traits. *Artificial Intelligence*, 15, 49–74.
- Carofiglio, V. & de Rosi, F. (2005). In favour of cognitive models of emotions. In *Proc. of the Joint Symposium on Virtual Social Agents – Mind-Making Agents, AISB 2005 Convention* (pp. 171–176), Hatfield, Hertfordshire, UK, April 12–15, 2005. AISB Press. www.aisb.org.uk/publications/proceedings/aisb05/10_Virt_Final.pdf
- Castelfranchi, C. (2000). Affective Appraisal Versus Cognitive Evaluation in Social Emotions and Interactions. In A. Paiva (Ed.) *Affective Interactions* (pp. 76–106) [LNAI 1814]. Berlin & Heidelberg: Springer-Verlag.
- Cavalluzzi, D., De Carolis, B., Carofiglio, V. & Grassano, G. (2003). Emotional dialogs with an embodied agent. In P. Brusilovsky, A. Corbett & F. de Rosi (Eds.), *User Modelling'03* (pp. 88–95) [LNAI 2702]. Berlin & Heidelberg: Springer-Verlag.
- Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16, 555–575.
- Conati, C. & MacLaren, H. (2005). Data-driven refinement of a probabilistic model of user affect. In L. Ardissono, P. Brna & A. Mitrovic (Eds.), *User Modeling 2005* (pp. 40–49) [LNAI 3538]. Berlin & Heidelberg: Springer.
- Datcu, D. & Rothkrantz, L. J. M. (2004). Automatic recognition of facial expressions using Bayesian belief networks. In *Proc. IEEE International Conference on Systems, Man and Cybernetics, IEEE SMC 2004* (pp. 2209–2214), October 10–13, 2004. The Hague, The Netherlands. IEEE Press.

- De Carolis, B., Pelachaud, C., Poggi, I. & de Rosis, F. (2001). Behavior planning for a reflexive agent. In *Proc. 17th Intl. Joint Conference on Artificial Intelligence (IJCAI 2001)* (pp. 1059–1066). Seattle, Washington, USA, August 4–10, 2001.
- De Rosis, F., Pelachaud, C., Poggi, I., De Carolis, N. & Carofiglio, V. (2003). From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *Intl. Journal of Human-Computer Studies*, 59 (1/2), 81–118.
- Elliott, C. & Siegle, G. (1993). Variables influencing the intensity of simulated affective states. In *Reasoning about Mental States – Formal Theories & Applications. Papes from the 1993 AAAI Spring Symposium* (pp. 58–67) [Technical Report SS-93-05]. Menlo Park, CA: AAAI Press.
- Forgas, J. P. (Ed.) (2000). *Feeling and thinking: The role of affect in social cognition*. Cambridge University Press.
- Gmytrasiewicz, P. & Lisetti, C. (2000). Using decision theory to formalize emotions. In S. Parsons & P. Gmytrasiewicz (Eds.), *Proceedings of the ICMAS-2000 Workshop on Decision Theoretic and Game Theoretic Agents* (pp. 39–47). Boston, MA, USA, July 10–12, 2000.
- Gratch, J. & Marsella, S. (2001). Tears and Fears: Modelling emotions and emotional behaviors in synthetic agents. In *Proceedings of the 5th International Conference on Autonomous Agents* (pp. 278–285). New York: ACM Press.
- Guojiang, W., Zhiliang W., Shadong, T., Yinggang, X. & Yujie, W. (2007). Emotion model of interactive virtual humans on the basis of MDP. *Frontiers of Electric and Electronic Engineering in China*, 2(2), 156–160.
- Hernandez, Y., Noguez, J., Sucar, E. & Arroyo-Figueroa, G. (2005). A probabilistic model of affective behavior for intelligent tutoring systems. In E. C. Segura & R. Whitty (Eds.), *MICAI 2005: Advances in Artificial Intelligence* (pp. 1175–1184) [LNCS 3789]. Berlin & Heidelberg: Springer.
- Hudlicka, E. (2003). Modeling Effects of Behavior Moderators on Performance: Evaluation of the MAMID Methodology and Architecture. In *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation*, Scottsdale, Az, May 2003.
- Ji, Q., Gray, W. D., Guhe, M. & Schoelles, M. J. (2004). Towards an integrated cognitive architecture for modeling and recognizing user affect. In E. Hudlicka & L. Cañamero (Eds.), *Architectures for Modeling Emotion: Cross-Disciplinary Foundations, Papers from the 2004 AAAI Spring Symposium* (pp. 77–78) [Technical Report SS-04-02]. Menlo Park, CA: AAAI Press.
- Li, X. & Ji, Q. (2005). Active affective state detection and user assistance with dynamic Bayesian networks. *IEEE Trans. Systems, Man and Cybernetics*, 35(1), 93–105.
- Miceli, M. (1992). How to make someone feel guilty. Strategies of guilt inducement and their goals. *Journal of the Theory of Social Behaviour*, 22, 81–104.
- Nicholson, A. E. & Brady, J. M. (1994). Dynamic Belief Networks for discrete monitoring. *IEEE Transactions on Systems, Man and Cybernetics*, 24(1), 1593–1610.
- Oatley, K. & Johnson-Laird, P. N. (1987). Towards a Cognitive Theory of Emotions. *Cognition and Emotion*, 1, 29–50.
- Ortony, A. (1988). Subjective importance and computational models of emotions. In V. Hamilton, G. H. Bower & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 321–333). Kluwer Academic Press.
- Ortony, A., Clore, G. L. & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.

- Pearl, A. J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufman Publishers.
- Picard, R. (1997). *Affective Computing*. Cambridge, MA: The MIT Press.
- Prendinger, H., Descamps, S. & Ishizuka, M. (2002). Scripting affective communication with life-like characters. *Applied Artificial Intelligence* [Special Issue on 'Merging cognition and affect in HCI'], 16 (7–8), 519–553.
- Prendinger, H. & Ishizuka, M. (2005). The empathic companion. A character-based interface that addresses users' affective states. *Applied Artificial Intelligence*, 19, 267–285.
- Sebe, N., Cohen, I. & Huang, T. S. (2005). Multimodal emotion recognition. In C. H. Chen & P. S. P. Wang (Eds.), *Handbook of Pattern Recognition and Computer Vision*. World Scientific.
- Sillince, J. A. A. & Minors, R. H. (1991). What makes a strong argument? Emotions, highly-placed values and role-playing. *Communication and Cognition*, 24 (3–4), 281–298.
- Staller, A & Petta, P. (2001). Introducing emotions into the computational study of social norms: a first evaluation. *Journal of Artificial Societies and Social Simulation*, 4 (1). <http://www.soc.surrey.ac.uk/JASSS/4/1/2.html>
- Walker, M. A., Cahn, J. & Whittaker, S. (1997). Improvising linguistic style. Social and affective bases for agent personality. In J. Lewi & B. Hayes-Roth (Eds.), *Proceedings of the 1st International Conference on Autonomous Agents* (pp. 96–105). New York: ACM Press.
- Zukerman, I. & Albrecht, D. W. (2001). Predictive statistical models for user modelling. *User Modelling and User-Adapted Interaction*, 11, 5–18.

Exercises in style for virtual humans

Zsófia Ruttkay, Catherine Pelachaud, Isabella Poggi
and Han Noot

1. Introduction

The title of this paper is inspired by Raymond Queneau's famous book *Exercises in Style* (Queneau 1981). In this ingenious literary work, the French author takes a banal story of a few lines set in a crowded bus, and tells it in different styles in 99 exercises. He does it so well that the reader can see the character acting before his eyes: how he gestures, whether he has a happy face or one of a bitter grumbler.

Another example of the power of style is *Creature comforts* (Aardman 1989), an Oscar-winning animation film in which animals talk and gesture in the easily-recognizable style of some human groups (of certain nationality and social status).

Style is thus a source of information on the speaker, as well as of variety and joy (or annoyance) when communicating with real people (Efron 1972; Knapp *et al.* 1997). Moreover, a pioneering empirical experiment has shown that such factors as the ethnicity and the personality (introvert/extravert) of a synthetic character – even if manifested in a simple, static feature – do have consequences on the effect of the character on the user (Nass *et al.* 2000).

One would like to benefit from style as well when confronted with virtual humans, also called embodied conversational agents (ECAs) (Cassell *et al.* 2000) in computer applications. Even if one should not expect a virtual human to act like a flesh-and-blood real person, the current situation of rather mechanical virtual characters devoid of style should be improved, even if only step by step.

Style is manifested in the usage of verbal and nonverbal signals to express some communicative function (Kendon 1993; McNeill 1992; Poggi 2003). Following the taxonomy of Poggi (Poggi 2007), gesture and gaze signals can bear meanings about the world (location and properties of objects, concepts or events, like “here”, “small” or “struggle”), about the speaker's affective state (such as “surprised”, “angry”) and meta-cognitive state (“I am concentrating”, “I am trying to remember”), beliefs (certainty, implausibility) and intentions (“I implore”, “I want to speak”, “I emphasize”, “I greet”), etc. A meaning may be conveyed by one or

more gestures. The mapping is often many-to-many: a meaning may be expressed by different gestures, and the same gesture may convey different meanings. For example, a beat may indicate a syntactical structure in the speech (e.g. enumeration), but also emphasis. In our discussion we use the notion of *gesture* for some motion of one or more body parts that has the goal of communicating some information.

1.1 Related work

In the world of traditionally-based (Thomas *et al.* 1981), as well as of computer-based animation it has been recognized how important it is to “add style” to (captured or synthesized) motion. The first steps have been taken in the direction of expressive ECAs, by endowing them with the capability of showing emotions (Gratch *et al.* 2001). Subtle issues like the impact of social role (Poggi *et al.* 2001; Prendinger *et al.* 2001; Niewiadomski & Pelachaud 2007) and personality (Ball *et al.* 2000) have been addressed. Also non-verbal signals have been used to accompany speech to make a virtual human more expressive and believable (Cassell *et al.* 1999; Lundeberg *et al.* 1999; Gratch & Marsella 2004; Kopp and Wachsmuth 2004; Pelachaud 2005; Heylen 2006; Gratch *et al.* 2007). However, these approaches concentrate on modeling the psychological, social and communicative aspects of the emotional and cognitive state. Usually the presentational issues (that is, how to visualize the emotional state) are not dealt with as a research topic, but as a practical task for an animator, often only to make a specific application or demonstrator.

A little work has addressed creative expressive ECAs on the signal level. Badler and his colleagues have developed EMOTE (Chi *et al.* 2000), a computational framework to modify expressiveness of hand, body and face gestures of ECAs (Byun *et al.* 2002). The Improv system (Perlin *et al.* 1996) allows one to modulate a given animation using a script language. While this system creates different types of behaviors, it does not embed any notion of style. Perlin (1995) demonstrated the importance of non-repetitiveness, by using some random selection criteria and noise to generate different instances of face and body motion of the character. Hartmann *et al.* (2006) view gesture expressivity as the qualitative value of a gesture execution and model it via 6 dimensions.

There have been initiatives to develop XML-based markup languages such as MPML (Tsutsui *et al.* 2000), VHML (Marriott *et al.* 2001), APML (De Carolis *et al.* 2002), RRL (Piwek *et al.* 2002), CML and AML (Arafa *et al.* 2002), MURML (Krandstedt *et al.* 2002) to encode some of the “human” aspects of multi-modal communication. Each of these representation languages acts either at the discourse and communicative functions level (APML, RRL, CML, MURML) or at

the signal level (AML, VHML). In each case the semantics of the control tags are given implicitly, expressed in terms of parameters (MPEG-4 FAP or BAP, muscle contraction, joint angles and the like) used for generating the animation of the expressive facial or hand gestures. They provide a direct link between what is often called the *mind* and the *body* of the agent. Lately, international effort has gathered to develop a unified ECA framework, called SAIBA where two representation languages are defined: FML (Function Markup Language) and BML (Behavior Markup Language) (Vilhjálmsón *et al.* 2007).

Until now none of the mark-up languages has addressed the style of the virtual human. In our view, style is a necessary additional level to be introduced to connect the communicative and the signal layer, allowing the explicit definition and manipulation of the mapping from the communicative tags to the signal tags. Our aim is on the one hand to develop such a style language, and in order to do so, we need to define those aspects that constitute style and describe them in terms of parameters. On the other hand we need to provide a computational model to use these style parameters, which can be embedded into the process of producing the final animation of a styled virtual human.

1.2 Our objectives

We are interested in creating ECAs that are expressive and individual in their behavior and appearance. In particular, we wish to endow ECAs with style in their non-verbal communicative behaviors. The complexity of the problem is manifested, for example, in the possibility of using different styles in the communicative act of greeting. Factors like culture, gender, age, personality, physical state and mood of the speaker, as well as characteristics of the situation (level of noise/visibility, characteristics of the listener) all contribute to decide if the greeting will be verbal and/or non-verbal, what facial expression and/or which hand will be used and in what way to express the greeting. Moreover, the different “sources of style” often describe conflicting behaviors: in certain social situations the style code may be different from the one a person follows normally. Personality and several other factors are also involved in how these conflicts are resolved. Different aspects of style have been studied independently, such as how context may affect the expression of emotion (Ekman 1999), how personality is perceived through behavior (Gallagher 1992), and what the dimensions of expressive gestures are (Laban 1974), but little is known of how style in its totality is manifested in the choice and characteristics of the gestures to be used. Therefore, the introduction of style requires a careful look at the following problems:

- Identifying aspects and parameters of style in human to human interaction.
- Providing a model to deal with conflicting style codes, as well as with the dynamical effects of the situation.
- Using these findings to define a language for style of ECAs.
- Identifying characteristics of gesturing and providing appropriate parameters to generate gestures which manifest certain aspects of the style.

We are proposing a hierarchical representation language that embeds all the above aspects for creating multi-modal styled ECAs. This language bridges the gap between our earlier work on the highest and lowest levels of nonverbal communication. In (Pelachaud *et al.* 2002) we proposed a framework that generates multi-modal dialogs: the dialog is marked with tags at the discourse level which then get instantiated as communicative functions when the conversational and social context is considered. In (Ruttkay *et al.* 2001) we have defined a constraint-based framework which allows the conceptual definition and on-the-fly generation of variants of facial expressions. In this paper we address how the higher level information on the character (such as culture, age) and on the situation (such as characteristics of the listener, physical circumstances in the environment) affect the choice and the performance of behaviors. We do not claim that we provide a complete and perfect model, as to do so would require modeling extremely complex factors such as culture, profession, gender, age...; how these factors each affect individually the behavior; what the combined effect is of several factors. Rather we are providing a framework as a first step in this direction, which allows us to test ideas. That is, we aim at developing a system with which one is able to create different looking animations corresponding to a set of factors, but we do not aim at modeling each of these factors. Thus we are interested in systematically visualizing various styles rather than modeling the process that gives rise to such a variation in the behavior. This is left for future work.

In Section 2 of this paper, we show how style is manifested in non-verbal communication among humans. In Section 3 we give the basic structure and parameters of the hierarchical GESTYLE language, and outline how they are used in the entire process of generating styled communication. Finally, we discuss the status of our work, and outline further research tasks.

2. What makes the human style?

For our investigation, we define *style* as the stable tendency in choosing non-verbal signals and their performance characteristics in communication, namely using facial, hand and body gestures, which accompany (and sometimes replace)

speech. Some examples: X has the tendency to be “stiff” and limited in facial and hand gestures, also in private informal situations; Y’s discourse is always accompanied by expressive gestures; Z uses the left-hand rather than the right-hand in gesturing; W avoids the other’s gaze.

Our basic hypothesis is that any communicative behavior, as well as any non communicative behavior, is aimed at some goal, whether biologically innate or culturally transmitted, whether conscious or unconscious (Poggi *et al.* 2000). For instance, one may eat a whole box of cookies to calm one’s hunger (conscious goal) or to calm one’s anxiety (unconscious goal). One may scream in fear because of an innate tendency to express a basic emotion (biological communicative goal), while one may smile hypocritically at a disliked person, out of politeness rules (cultural communicative goal). The combination and interaction of multiple goals that motivate the person are manifested in one’s style. This manifestation takes place in two layers: the ultimate goals of the person (e.g. being noticed and acknowledged by others) are translated into general *semantic characteristics* of their nonverbal communication (e.g. preference for frequent and visible gestures), which then are manifested in the syntactical and *morphological details* of the final animation of the body (e.g. hand gesturing with big amplitudes).

2.1 Permanent and contingent goals behind the style

In every-day communication, two kinds of factors affect the final outcome of a person’s communication: permanent and contingent ones. The former are the goals and resources coming from the biological and cultural endowment, that are always active; the latter are the goals activated and the resources provided by the contingent situation in which one has to communicate.

Our *permanent goals* of survival and reproduction generate some subgoals and sub-subgoals that are presumably innate and universally shared in the human: we all have the goal of physical safety and of being loved, of being well thought-of by others and by ourselves, of not being too greatly dependent on other people and of living up to our own self; also, we generally care for other people’s well-being, as witnessed by the existence of feelings like tenderness, the sense of justice, and the emotions of guilt and compassion. All these goals, in our view, continuously impinge on us and may be dismissed only in pathological conditions or when they conflict with each other. In order to reach these permanent goals we are endowed with a set of *long-lasting resources* that are our internal characteristics and capacities, such as personality and cognitive traits, gender, age, and cultural roots.

Among a person's cognitive resources, some are probably innate, like a higher or lower capacity to make inferences, or the different aptitudes towards visual, linguistic or other skills. These resources and aptitudes, when activated, do have a visible manifestation subject to style, for instance in preferences for making spatial illustrations with hands to accompany speech, or in the effort of thinking, indicated by the tempo of speech, the facial expressions (forehead wrinkling, eyebrow squeeze) and the head or hand gestures (hand(s) on the forehead or covering the eyes). But a number of other cognitive capacities are culturally learned (Argyle 1975; Morris *et al.* 1979; Bourdieu 1998), namely knowledge of the world, the communication repertoires (verbal and nonverbal) one comes to learn from infancy on, and generally cultural knowledge. According to de Rosis and colleagues (2004), culture also provides a set of norms on how to do things, how to behave, what and how to communicate; and these norms are goals that are part of the permanent endowment of a person's multiple goals when he/she starts a communication.

The major *contingent goals* are the intentions to express the specific meaning the person wishes to communicate. The contingent resources of the speaker and listener are (Laban & Lawrence 1974; Carbonell 1980): *motor energy* (only for the speaker), *available modalities*, current level of *communicative and cognitive capacity* and *emotional state* of both the speaker and the listener(s), and *personality* and *culture* of the current listener(s). The resources related to the situation are time and availability and quality of communication channels. For instance, a tired person will speak quietly and will avoid conspicuous gestures, but if the listener has difficulty with hearing, even a tired person will talk more loudly, and may even choose to use gestures in face to face situations. The *social setting* is relevant too: we use more polite words and fewer gestures when talking to a high status person or to someone in a formal situation, and fewer colloquial expressions and gestures in public than in private. Finally, there may be characteristics of a person's communication which are totally *idiosyncratic* and cannot be derived from the above mentioned norms. For example, the preference for using a hand, or the way of performing a common gesture can be typical of an individual.

During communication, the internal and external transitory conditions imply temporary new goals extending the permanent ones, and impose resource constraints. Each of the goals determines the choice of a modality or set of modalities to be used, and/or the choice of a particular signal or set of signals, and/or the characteristics of the production of signals. All the goals and constraints which are valid in a given moment contribute to the final output. We find it important to relate factors involved in style characteristics to goals, as a basic standpoint, though providing a complete computational model for the mapping of goals to semantic characteristics is beyond the scope of this paper and our present capabilities.

Our work on style differs from others, like (Walker *et al.* 1997), mainly in three respects. First, we consider a wider set of factors that contribute to style, not only the degree of formality and power relationships. Second, our notion of style concerns also the permanent goals of the individual, not only the contingent ones. Finally, we are interested in style differences in the non linguistic aspects of communicative behavior.

2.2 Style manifested in gesture characteristics

The goals explained above are manifested in the gestures used. The gesticulation of a person can be described along four dimensions: level of redundancy, threshold for using a gesture, the repertoire of the gestures used, and the motion characteristics of the gestures used (Argyle 1975; Knapp *et al.* 1997; Morris *et al.* 1979).

Redundancy

Some people use gestures only as a substitute for words, others also when the gesture simply repeats the meaning conveyed in the speech. Possible motivations for redundancy are the low cognitive capacity of the listener, or the noise in the verbal and/or visual channels, or the high motivation of the speaker to be understood/noticed.

Repertoire

A repertoire is the set of gestures used by a person. For a single meaning, different alternative gestures may be used. Some people have a richer repertoire of gestures than others. This is due partly to cultural factors (the repertoire of symbolic gestures in Italy, for instance, is wider than in Great Britain), partly to personality and partly to other individual factors.

Threshold

Two people may have the same gesture repertoire, but they may judge the appropriateness of a gesture in a given situation differently. The above mentioned characteristics (personality, emotional state, familiarity with the listener) can have a range of (discrete) intensities, and a threshold can be given to use certain gestures.

Motion characteristics

Last, but not least, there is a stylistic difference in motion characteristics, that is, the way of performing gestures. A hand gesture can be analyzed in terms of some formational parameters (Stokoe 1978), like hand shape (form of the hand in

making it), location (where the hand moves), orientation (direction of palm and metacarp); and movement parameters like speed and smoothness. While some parameters are determined by lexical and sociolinguistic variation (that is, they assure that the signal is recognized as one with a specific meaning), others, determining the final movement, are free to manifest the style. Specifically, while the hand shape or the direction of movement are restricted according to the gestural dialect of some culture or region, the gestural style of a person may be expressed in variations of the following characteristics of the motion (Laban & Lawrence 1974; Chi *et al.* 2000; Hartmann *et al.* 2004):

- *Tension*: movement being tense or relaxed, where tension is manifested in the change of speed of the motion;
- *Amplitude*: how big the motion is, it can be wide or narrow;
- *Manner*: the shape of the path of the hand can be smooth or angular;
- *Tempo*: slow or fast.

3. The GESTYLE language

For designing and controlling the gesturing behavior of virtual humans, we designed a language that can express all the factors contributing to the non-verbal style as described in the previous section. This is done by using three kinds of markups, each of which define some of the above introduced factors of style. The fourth type of markup indicates the meaning to be expressed, possibly also by some styled gestures (we assume that the virtual character is to speak a text, though gestures may be used without textual correspondent). Hence, the first three types of markups, of different time scopes and origins, will decide what gesture will be used, and in what form, to accompany or extend the speech. In the process of making the choices for the gesture parameters, conflicting gesture parameters may be prescribed, thus conflict resolution is to be dealt with too.

3.1 The hierarchy of four markup types

In order to model the above explained complex relation between static and dynamical parameters influencing the gesturing style, we have defined the GESTYLE language as a hierarchical language: some tags defined at a higher level may imply tags of a lower level. GESTYLE allows four types of markup tags:

The CHARACTER MARKUP tags define the virtual character's static characteristics, as culture (having values of ethnic group or sub-group of an ethnic group like "educated British" or "Neapolitan"), personality (e.g. extrovert / intro-

vert), social role (having value as profession and/or social status), age and sex to capture biological aspects, and eventual individual characteristics like handedness, special gesture usage. This information may be considered invariant during the period the agent is conversing.

The SITUATION MARKUP tags specify a situation, by setting dynamical aspects of the speaker (mood, physical state, momentarily available communicative modalities) and of the environment (characteristics of the addressee, the objects in the environment, etc.).

The COMMUNICATIVE MARKUP tags are used to annotate the text with information the agent desires to communicate (consciously or unconsciously); this information will (potentially) be accompanied by some gesture. The information may be about the world (characteristics of objects), the speaker's mental state (emotion), the discourse state (taking/giving turn), etc.

The GESTURE MARKUP tags prescribe a gesturing sequence, by specifying what gestures are to be expressed at certain points of time: raised eyebrow, head nod, wave right hand, etc. Some parameters may be given to the gestures, like amplitude, duration, start/end time, and motion-manner. Time parameters may be given qualitatively (short / long duration), or left partially undefined (start when an utterance starts; perform the gesture during a certain utterance). These relative times must then be expressed, also on the basis of timing information from a text-to-speech system or audio system (responsible for providing the speech to be accompanied by the gesturing) in terms of absolute times. In general, the GESTURE MARKUP tags are generated according to the above three classes of tags. However, it is also possible to insert GESTURE MARKUP tags explicitly into the text, in order to define characteristics of the gestures of a given modality (e.g. to make the motion of the right-hand slow) for some time interval, or even to overwrite or extend the generated gestures. This low-level direct influence makes it possible to prescribe some specific gesturing that cannot be captured by the high-level parameters. It can also be used as a direct scripting language to define and test animations.

At the lowest motion definition level (which is not supported by markup tags, but considered to be implemented by an interface to the animation engine used) the GESTURE MARKUP tags are expressed in terms of parameters (e.g. muscle contraction for facial physically-based model, joint angle for articulated body) which can be fed directly to the animation player. We are using MPEG-4 facial animation parameters (ISO 1998) and MPEG-4 body animation parameters with H-Anim standards (H-Anim 2002).

The GESTYLE language allows the handling of conflicts in three ways:

```

<characterdef id="char1" culture="dutch" sex="male" profession="engineer"
  handedness="left" weights="1 1 2 2">
<situation location="public-informal" emotion="sad">
  Yeah, I am in a bad mood as you can see. I got a parking ticket of 40 euro.
  <situation emotion="angry">
    You know, they ticketed the entire street, though the machine was out of
    order.
  </situation>
  Well, let's have <communicate locate="the beer glass in front of me"> another
  one of this. <communicate/>
  Ah... that tastes good.
  <situation physical-state="slightly-drunk" weight="3">
    <communicate locate="char2"> You know what? </communicate>I simply
    <communicate act="emphasize"> <gesture name="rude_gesture"
    duration="short"/> will not pay </communicate>the fee this time!
  </situation>
</situation>
</characterdef>

```

Figure 1. A piece of conversational text marked up with different GESTYLE tags

- A built-in default preference for the different high-level parameters (e.g. personal is preferred over professional, which is preferred over cultural);
- Giving the preferences explicitly, as part of the definition of the character's personality (e.g. a person of low self-esteem will confirm more to the social norms and to the norms of the listener);
- Allow any “black box” type reasoning module to decide in case of conflicts.

The parameters are given in an XML-compliant Markup Language. Its tags are used to annotate the text to be spoken by the character. In Figure 1, a piece of text is given annotated with all 4 types of tags.

In general, communicative gestures accompany speech, and have to be synchronized with speech. The actual frequency and timing of some repetitive gestures (e.g. blink) can be generated on the basis of biological characteristics, while others (e.g. idle motion) may be defined on the basis of personality characteristics (e.g. if the character is nervous, he will make more frequent idle motions). The interplay of the hierarchy of these language aspects is illustrated in Figure 2. Note that all the tags at all levels, in principle, may be either computed by a dialog generator (Pelachaud *et al.* 2002) or may be specified manually. This allows the language to be used by applications and/or as a scripting language for animators on different levels of detail.

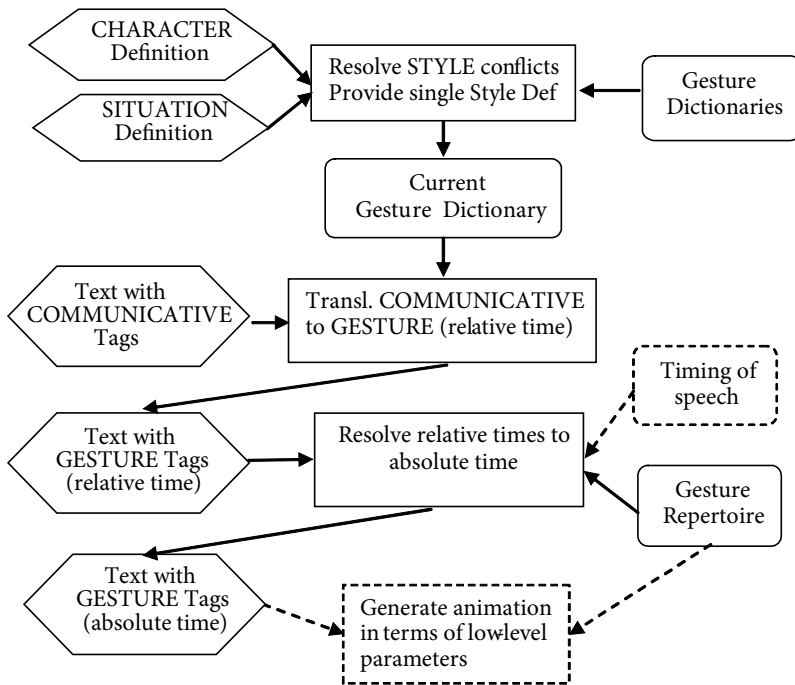


Figure 2. The simplest usage of the four types of markups in the GESTYLE language, when the SITUATION is given once, and the GESTURE tags are derived from the COMMUNICATIVE tags. It is also possible to process text with multiple SITUATION definitions and given GESTURE markup tags.

3.2 The language of composite gestures

We have developed a language that describes gestures as a combination of basic gestures. A basic gesture may be viewed as the minimum action a body part can carry out (look_left, raise_left_eyebrow). Two temporal operators have been defined: ‘+’ for parallel combination and ‘*’ for sequential concatenation. Gestures may be created by combining basic gestures with these operators. As the operators can be used in a nested way, a wide variety of gestures can be defined. For instance the facial gesture corresponding to “surprise” may be composed of the basic facial gestures: raised eyebrow + raised upper lid + open mouth. The gesture of nodding is a sequence of nods, while a single nod is the sequence of three elementary motions of head_down * hold * head_straight.

The formulation of basic hand gesture is based on a functional separation of arm position, wrist orientation and hand shape, using MPEG-4 (ISO 1998) or

H-Anim (H-Anim 2002) coding systems. For facial gestures, we use the common definition of emotions (Ekman 1999) and our own earlier work on cognitive facial expressions (Poggi *et al.* 2000).

3.3 The gesture dictionaries

It is an essential characteristic of an ECA what gestures it uses to express meanings. Ideally, we would like to derive the gestures used from goals and motivations, but presently we lack the essential psychological and sociological knowledge to act as a base for a reasoning system as a computational model, covering all aspects and resolving conflicts. As an alternative, and also to develop experiments and test hypothesis, we use a dictionary-based probabilistic approach.

In the gesture dictionary, a set of alternatives of (uni- or multimodal) gestures are given for each meaning. We model the characteristics of usage of gestures marking the same communicative act by assigning probabilities to the individual gestures. Taking this characteristic into account too, the gestures used for a communicative act are given by a gesture dictionary entry of the following form:

communicative_act (*parameters*₁, *Gesture*₁, *P*₁), ... (*parameters*_n, *Gesture*_n, *P*_n)

where *Gesture*₁, ..., *Gesture*_n are gestures, covering the alternatives of expressing the communicative function, and *P*₁, ..., *P*_n are probabilities of using the specific gesture in this role.

The optional gesture-modifying parameters specify the motion characteristics of the gesture. Different gesture dictionaries can be defined and given to set different “gesturing codes”: one for gesturing habits of a culture (e.g. Italian gesturing) or profession (e.g. teacher-like gesturing), one for special gesturing of a certain person.

3.4 Selection of a gesture to be used

When making an ECA gesture stylishly, a single gesture has to be selected from the possible alternatives, prescribed by different gesture dictionaries according to different CHARACTER specification parameters of the character. The choice may be further influenced by the current SITUATION parameters. For each moment, a current gesture dictionary is composed from the ones prescribed in the CHARACTER and SITUATION definition (see below). In the style definition multiple gesture dictionaries are referred to, some with a weight factor and others in a strict hierarchy. From these a single current dictionary is compiled by first including all communicative acts which only occur in strictly one of the source

dictionaries. Then for conflicting prescriptions, that is meaning entries which occur in more than one dictionary, the hierarchy or the explicitly given weights are used to select one.

3.5 Instantiation of styled gesture

When instantiated, a basic gesture corresponds to an animation of the facial features or body parts involved. A gesture may be instantiated in two ways which vary in complexity. In the simplest case, instantiation of a gesture takes place by instantiating the basic gestures it is composed of. In a more sophisticated framework, a gesture is defined as a set of basic gestures linked by constraints, expressing e.g. flexible timing or asymmetry conditions (Ruttkay *et al.* 2001).

Some characteristics of the gestures, like duration, intensity, onset/offset time (for facial gestures), preparation/hold/withdrawal time (for hand gestures) can be specified. This framework enables the modification of a gesture definition to produce variants of it with different *amplitude*, *motion manner* or *tempo*. Such a definition of gesture increases the flexibility of the creation process for animation as well as allowing the non-repetition of the final animation for each gesture instantiation.

In the *gesture generation stage*, all the given CHARACTER, SITUATION and GESTURE tags having an effect on the motion of a gesture which has to be produced will be taken into account. For example, an introvert person will make less articulated gestures, while a typical asymmetric eyebrow-usage will have an effect on all facial signals involving eyebrows. The effect of high-level CHARACTER and SITUATION tags on the motion characteristics are given in terms of low-level GESTURE tags. Possible conflicts are dealt with in a similar way as in the gesture selection stage. When nesting occurs, the locally-defined, deepest GESTURE parameters override temporarily all other prescriptions.

The on-the-fly generation of individual expressive gestures is based on our earlier work on the *gesture repertoire principle*: a gesture is defined in terms of characteristics of the shape and motion of the involved features (Ruttkay 2001). When GESTURE parameters are to be applied to the “standard” definition of a gesture, they are expressed in terms of modifying certain constraints. For instance, the amplitude increase of the smile will be expressed as an increase in the extreme positions. However, as in the definition of smile the limitations of application/release speed are incorporated, the increase of amplitude may result in increase in duration too. The constraint framework allows the generation of different instances of a gesture, including random variants.

It may happen that some modalities must be used for different gestures at the same time (as in the case of speech and smile). For these modalities, a blend of the contributing gestures is to be produced, either as a weighted sum of the contributing gesture parameter functions, or in a more sophisticated way, taking into account the constraints that should hold for the gestures.

4. Conclusion and further work

Until recently, we have been working on generating facial expressions. We had developed a conversational agent (Pelachaud *et al.* 2002) and an interactive editor to specify facial expressions (Noot *et al.* 2000), also in terms of constraints. We have developed two systems for defining and animating hand gestures (Hartmann *et al.* 2002; Ruttkay *et al.* 2003). The GESTYLE language has been implemented (Noot *et al.* 2003). We have also developed a reasoning-based system (Pelachaud & Poggi 2002), covering some aspects of style, and acting on the goals and norms. In deriving which gesture to use, possible modality conflicts are also taken into account.

Using these tools and taxonomy of facial and hand gestures, we are currently implementing ECAs capable of styled expressions using the modalities of face, head and hands. We will experiment with the effect of variations and non-determinism within gestures. Later on, we wish to test if the style of synthetic characters, manifested in facial and hand gestures is perceived as intended.

There are several issues which need to be clarified before starting to implement a more full-fledged system. First of all, different dictionaries need to be defined from psychological and sociological studies to find out what the distinctive gesture types and manners are for certain groups (divided by culture, profession, age and so on). It is also a question if the gesturing of cultures could be defined in terms of more refined concepts, like every-day and social values, living conditions.

Ultimately, one would like to have an ECA which manifests style also in the verbal modality. A markup language for emotional speech has been developed, which fits within the framework of our GESTYLE markup languages (Van Moppes 2002). It is possible to define speech style: the person- or culture-dependent way of expressing emotions, emphasis, hesitation, etc. However, style is manifested very strongly through choice of words and sentence structures. There is ongoing work in generating styled natural language content (Walker *et al.* 1997). In the longer term, it is a challenging task to develop ECAs which have a consistent style in all their modalities.

Acknowledgements

We are thankful to Fiorella de Rosis for her constant advice, to Massimo Bilvi and Elisabetta Bevacqua for implementing the Greta system, and to Anton Eliëns and Zhisheng Huang for making their STEP system available for experimenting with hand gestures.

References

- Aardman Studios (1989). *Creature Comforts*. <http://www.aardman.com/>.
- Arafa, Y., Kamyab, K., Kshirsagar, S., Guye-Vuilleme, A. & Thalmann, N. (2002). Two Approaches to Scripting Character Animation. In *Proc. of the AAMAS Workshop on Embodied conversational agents – Let's specify and evaluate them!*, Bologna, Italy, July 16, 2002.
- Argyle, M. (1975). *Bodily Communication*. London: Methuen and Co. Ltd.
- Ball, G. & Breese, J. (2000). Emotion and personality in a conversational agent. In Cassell *et al.* 2000 (pp. 189–219).
- Bourdieu, P. (1998). *Practical reason: On the theory of action*. Stanford University Press.
- Byun, M. & Badler, N. (2002). FacEMOTE: Qualitative parametric modifiers for facial animations. In *Proc. of the ACM SIGGRAPH Symposium on Computer Animation* (pp. 65–72). San Antonio, TX, July 21–22, 2002.
- Carbonell, J. G. (1980). Towards a process model of human personality traits, *Journal of Artificial Intelligence*, 15, 49–74.
- Cassell J., Sullivan J., Prevost S. & Churchill E. (Eds.) (2000). *Embodied Conversational Agents*. Cambridge, MA: MIT Press.
- Cassell, J., Bickmore, T., Billingham, M., Campbell, L., Chang, K., Vilhjálmsson, H. & Yan, H. (1999). Embodiment in Conversational Interfaces: Rea. In *ACM CHI'99 Conference Proceedings* (pp. 520–527). Pittsburgh, PA, May 15–20, 1999.
- Chi, D., Costa, M., Zhao, L. & Badler, N. (2000). The EMOTE Model for Effort and Shape. In *Proc. of SIGGRAPH 2000* (pp. 173–182). New Orleans, Louisiana, July 23–28, 2000.
- de Carolis, Carofiglio, Bilvi, M. & Pelachaud, C. (2002). APMML, a Mark-up Language for Believable Behavior Generation. In *Proc. of the AAMAS Workshop on Embodied conversational agents – Let's specify and evaluate them!*. Bologna, Italy, July 16, 2002.
- de Rosis, F., Pelachaud, C., Poggi, I., De Carolis, N. & Carofiglio, V. (2003). From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59, 81–118.
- de Rosis, F., Pelachaud, C. & Poggi, I. (2004). Transcultural believability in embodied agents: A matter of consistent adaptation. In R. Trappl & S. Payr (Eds.), *Agent Culture: Designing virtual characters for a multi-cultural world* (pp. 78–106). Dordrecht: Kluwer Academic Publishers.
- Efron, D. (1972). *Gesture, Race, and Culture*. The Hague: Mouton.
- Ekman, P. (1999). Facial Expressions. In T. Dalgleish & Power (Eds.), *The Handbook of Cognition and Emotion* (pp. 301–320). John Wiley & Sons, Ltd.
- Gallagher, P. E. (1992). Individual differences in nonverbal behavior: Dimensions of style. *Journal of Personality and Social Psychology*, 63(1), 133–145.

- Gratch, J. & S. Marsella (2001). Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proc. of Fifth Intl. Conference on Autonomous Agents* (pp. 2778–285). ACM Press.
- Gratch, J. & Marsella S. (2004). A domain-independent Framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4), 269–306.
- Gratch, J., Wang, N., Gerten, J., Fast, E. & Duffy, R. (2007). Creating Rapport with Virtual Agents. In C. Pelachaud, J.-C. Martin, 7th E. André, G. Chollet, K. Karpouzis & D. Pelé (Eds.), *Intelligent Virtual Agents: 7th International Working Conference, IVA 2007* (pp. 125–138) [LNAI 4722]. Berlin & Heidelberg: Springer.
- H-anim (2002). Humanoid animation working group. <http://www.h-anim.org/>
- Hartmann, B., Mancini, M. & Pelachaud, C. (2002). Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis. In *Proc. of Computer Animation 2002* (pp. 111–119). IEEE Computer Society Press.
- Hartmann, B., Mancini, M. & Pelachaud, C. (2004). Expressivity Control for Multimodal Behavior Synthesis. In *Poster Proceedings of ACM SIGGRAPH / Eurographics Symposium on Computer Animation* (pp. 24–25). Grenoble, France, August 27–29, 2004.
- Heylen, D. (2006). Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics*, 3(3), 241–267.
- ISO (1998). *Information Technology – Generic coding of audio-visual objects – Part 2: visual, ISO/IEC 14496-2 Final Draft International Standard*, Atlantic City.
- Kendon, A. (1993). Human gesture. In Ingold T. & Gibson K. (Eds.), *Tools, Language and Intelligence* (pp. 43–62). Cambridge University Press.
- Knapp, M. L. & Hall, J. A. (1997). *Nonverbal communication in human interaction*. Harcourt Brace.
- Kopp, S. & Wachsmuth, I. (2004). Synthesizing Multimodal Utterances for Conversational Agents. *Computer Animation and Virtual Worlds*, 15(1), 39–52.
- Krandsted, A., Kopp, S. & Wachsmuth, I. (2002). MURML: A multimodal utterance representation markup language for conversational agents. In *Proc. of the AAMAS Workshop on Embodied conversational agents – Let's specify and evaluate them!* Bologna, Italy, July 16, 2002.
- Laban, R. & Lawrence, F. C. (1974). *Effort: Economy in body movement*. Boston: Plays, Inc.
- Lundeberg, M. & Beskow, J. (1999). Developing a 3D-agent for the August dialogue system. In *Proceedings of the ESCA Workshop on Audio-Visual Speech Processing*. Santa Cruz, CA, USA, August 7–9, 1999.
- Marriott, A., Beard, S., Stallo, J. & Huynh, Q. (2001). VHML – Directing a Talking Head. In *Proc. Sixth International Computer Science Conference* (pp. 18–20) [LNCS 2252]. Berlin & Heidelberg: Springer.
- McNeill, D. (1991). *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press.
- Morris, D., Collet, P., Marsh, P., & O'Shaughnessy, M. (1979). *Gestures. Their origins and distribution*. London.
- Nass, C., Isbister, K. & Lee, E.-J. (2000). Truth is beauty: Researching embodied conversational agents. In Cassell *et al.* 2000 (pp. 374–402).
- Niewiadomski, R. & Pelachaud, C. (2007). Model of Facial Expressions Management for an Embodied Conversational Agent, ACII, Lisbon, September.
- Noot, H. & Ruttkay, Z. (2000). CharToon 2.0 Manual, CWI Report INS-R0004, Amsterdam.

- Noot, H. & Ruttkay, Z. (2004). Style in Gesture. In A. Camurri & G. Volpe (Eds.), *5th International Gesture Workshop, GW 2003. Genova, Italy, April 15–17, 2003, Selected Revised Papers* (pp. 324–337) [LNCS 2915]. Berlin: Springer-Verlag
- Ortony, A., Clore, G. L. & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press,
- Pelachaud, C. (2005). Multimodal Expressive Embodied Conversational Agents. In H. Zhang, T.-S. Chua, R. Steinmetz, M. Kankanhalli & L. Wilcox (Eds.), *Proceedings of the 13th annual ACM international conference on Multimedia* (pp. 683–689). New York, NY: ACM Press.
- Pelachaud, C., Carofiglio, V., De Carolis, B. De Rosis, F. & Poggi, I. (2002). Embodied contextual agent in information delivering application. In *Proc. of AAMAS'04* (pp. 758–765). ACM press.
- Pelachaud, C. & Poggi, I. (2002). Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation*, 13, 301–312.
- Perlin, K. (1995). Real time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics*, 1(1), 5–15.
- Perlin, K. & Goldberg, A. (1996). Improv: A system for interactive actors in virtual worlds. In *Proc. of SIGGRAPH 1996* (pp. 205–216). New Orleans, Louisiana, August 4–9, 1996.
- Piwek, P., Krenn, B., Schröder, M., Grice, S., Baumann & Pirker, H. (2002). RRL: A Rich Representation Language for the Description of Agent Behaviour in NECA. In *Proceedings of the AAMAS workshop on Embodied conversational agents – let's specify and evaluate them!* Bologna, Italy, July 16, 2002.
- Poggi, I. (2007). *Mind, Hands, Face and Body. A goal and belief view of multimodal communication*. Berlin: Weidler Verlag.
- Poggi, I. (2007). Mind, Hands, Face and Body. A Goal and Belief View of Multimodal Communication. In H. Kalverkämper, R. Krüger & R. Posner (Eds.), *Körper, Zeichen, Kultur (Body, Sign, Culture) 19*. Berlin: Weidler Verlag.
- Poggi, I. (2003). Mind Markers. In C. Rector, I. Poggi & N. Trigo (Eds.), *Gestures: Meaning and Use* (pp. 119–132). Oporto: Universidad Fernando Pessoa Press.
- Poggi, I. & Pelachaud, C. (2000). Facial Performative in a Conversational System. In Cassell *et al.* 2000 (pp. 155–188).
- Poggi, I., Pelachaud, C. & de Carolis, B. (2001). To display or not to display? Towards the architecture of a Reflexive Agent. In *Proceedings of the 2nd Workshop on Attitude, Personality and Emotions in User-adapted Interaction*, held in conjunction with *User Modeling 2001*. Sonthofen, Germany, July 13–17, 2001.
- Prendinger, H. & Ishizuka, M. (2001). Social role awareness in animated agents. In *Proc. of Fifth Intl. Autonomous Agents Conference* (pp. 270–277). ACM Press.
- Queneau, R. (1981). *Exercises in Style*. New Directions (English translation by B. Wright).
- Ruttkay, Z. & Noot, H. (2001). FESINC: Facial Expression Sculpturing with Interval Constraints. In *Proc. of AA'01 Workshop Representing and Annotating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents*. Montreal, Canada, May 28–29, 2001.
- Ruttkay, Z. (2001). Constraint-based facial animation. *Intl. Journal of Constraints*, 6, 85–113.
- Ruttkay, Z., Huang, Z. & Eliëns, A. (2004). The conductor – Gestures for embodied agents with logic programming. In K. R. Apt, F. Fages, F. de Rossi, P. Szeredi, J. Váncza (Eds.), *Joint ERCIM/CoLogNET International Workshop on Constraint Solving and Constraint Logic Pro-*

- gramming, CSELP 2003, Budapest, Hungary, June 30 – July 2, 2003, *Selected Papers*. [LNAI 3010] (pp. 266–284). Berlin & Heidelberg: Springer-Verlag.
- Stokoe, W. C. (1978). *Sign language structure: An outline of the communicative systems of the American deaf*. Silver Spring: Linstock Press.
- Tsutsui, T. Saeyor, S. & Ishizuka, M. (2000). MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions. In (CD-ROM) *Proc. WebNet 2000 World Conf. on the WWW and Internet*. San Antonio, Texas, October 30 – November 4, 2000.
- Thomas, F. & Johnston, O. (1981). *Disney animation: The illusion of life*. New York: Abbeville Press.
- Van Moppes, V. (2002). *Improving the quality of synthesized speech through mark-up of input text with emotions*. Unpublished Master Thesis, Free University (VU), Amsterdam.
- Vilhjalmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., Ruttkay, Z., Thórisson, K. R., van Welbergen, H. & van der Werf, R. (2007). The Behavior Markup Language: Recent Developments and Challenges. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis & D. Pelé (Eds.), *Intelligent Virtual Agents: 7th International Working Conference, IVA 2007* (pp. 99–111) [LNAI 4722]. Berlin & Heidelberg: Springer.
- Walker, M., Cahn, J. & Whittaker, S. (1997). Improvising linguistic style: Social and affective bases for agent personality. In W. Lewis Johnson (Ed.), *Proceedings of the First International Conference on Autonomous Agents, Agents'97* (pp. 96–105). New York, NY: ACM Press.

Expressive characters in anti-bullying education

Ruth Aylett, Ana Paiva, Sarah Woods, Lynne Hall
and Carsten Zoll

1. Introduction

Expressive behaviour can be considered a subset of behaviour in general for an embodied agent, whether this is a living human, a robot or a graphically embodied synthetic character. We can think of it as behaviour that is interpreted by others as directly signalling an affective state. It may be a distinctly produced behaviour, as in facial expression or gesture, but it may also be a behavioural modifier as in posture or tone of voice. Because in principle it tells the observer something about the internal state of the agent, it acts as an additional communication mechanism alongside more externally directed and functional behaviour, whether physical actions or the use of natural language. Indeed it is argued that in the case of humans, around 80% of overall communication comes through these indirect channels (Argyle 1972).

Knowing something about the affective state of an agent is important to a communicative partner for a number of reasons. It is the basis of social relationship-building ('does this person like me or not?'); of the creation of trust ('Is this person telling me the truth') as well of the understanding of the motives and goals which help to fill in the context of the functional behaviour and allow predictive inferencing about what an agent will do next. This can help to produce a feeling of coherent action that is required for a human to feel that in some sense they 'understand' what an agent is doing. In this way, knowledge of affective state acts as a major support to the application of a 'theory of mind' (ToM) to other agents. Its importance is underlined by the human tendency to anthropomorphize the behaviour of non-human agents in terms of human affective states and motives, from pets to inanimate objects such as cars and other machines.

Because of this importance, expressive behaviour is seen as a very significant issue in the development of believable graphically-embodied characters, as can be seen from the other chapters of this volume. A difficult term to define, believability

(Bates 1994) is seen as the extent to which a human is willing to suspend his or her disbelief in a collection of graphical pixels and to see it as an autonomous entity with 'its own' internal life. Given that affective engagement with the world is such an important component of human internal life, it can be seen that appropriate expressive behaviour can make a major contribution to believability.

In this chapter we discuss the role of expressive behaviour in the graphically-embodied characters originally developed as part of the EU-funded project VICTEC – Virtual ICT with Empathic Characters (IST-2001-33310). This sought to apply virtual dramas acted out by 3D graphically-embodied characters to what is known generically in the UK as Personal and Social Education (PSE) (or more recently as Personal, Social and Health Education – PSHE). This covers topics such as education against bullying and racism, on drugs, including smoking and alcohol, and sex education. A common thread in these topics is that knowledge in and of itself is not sufficient to meet the pedagogical objectives, since attitudes and emotions are at least as important to producing desired rather than undesired behaviour. For this reason, techniques such as small-group discussion, role-play and dramatic performance by Theatre-in-Education (TiE) groups may be used. A motivation for the work was to try to create some of the impact of dramatic performance through virtual dramas. The specific topic selected was anti-bullying education. Effective though theatrical performance is in this domain, it is necessarily collective, and in any group it is very likely that some individuals will be victims of bullying by some other in the group and thus will be inhibited in their participation. Therefore, a virtual drama application that could be used by the individual seemed to have a possible use.

We first consider the characteristics of bullying and go on to describe the demonstrator FearNot! produced by the VICTEC project. We discuss empathy and its role in meeting the pedagogical objectives of FearNot! We will go on to argue that consistency along the main dimensions of empathy is more important than the degree of naturalistic fidelity of expressive behaviour and cite evaluation results supporting this position.

2. Bullying and FearNot!

Bullying in social relationships can be differentiated from merely aggressive behaviour by its frequency, planned nature and dependence on a perceived inequality of strength, power or status. In a widely accepted definition of bullying in schools:

A student is being bullied or victimised when he or she is exposed *repeatedly* and *over time* to *negative action* on the part of one or more other students.

(Olweus 1991)

Bullying may involve hitting, punching, blackmail, threats and spiteful behaviour, including overt abstraction or damage of possessions. This is known as *direct bullying*. *Verbal bullying* includes insults, mockery and the spreading of lies, whether verbally, by notes or more modern technology such as email or texting. A further category of bullying is known as *relational bullying*:

The purposeful damage and manipulation of peer relationships and feelings of social exclusion.

(Crick & Grotpeter 1995)

This includes for example refusing to sit next to someone, or excluding them from group activities whether socially or in study. Studies have shown that bullying is a widespread problem in schools – for example a study carried out in the UK and Germany in the period 1996–1998 (Wolke *et al.* 2001) showed that 24% of UK school students aged 9–11 said they were bullied every week, while between 12% and 17% of the sample admitted to bullying others at least 4 times in the previous term.

A characteristic of education against bullying is that there is no definite strategy that will always work. Even the action urged in the generally-agreed educational message “tell someone you trust, don’t suffer in silence” is not guaranteed to succeed. “Hit the bully back” is an example of controversial advice: often offered by parents, it is opposed by teachers. In addition it only succeeds in a minority of direct bullying cases, but because of its dramatic effect, success may well be over-reported. Thus the educational objectives of anti-bullying education, while partly intended to show that bullying is a ‘bad thing’ and to dramatise the effects upon the victim, are necessarily diverse. Equipping children with a greater understanding of social dynamics and demonstrating the range of coping strategies and the circumstances under which they succeed or fail is therefore a valid pedagogical approach.

The response discussed here is the development of a virtual theatre application called FearNot! (Fun with Empathic Agents to Reach Novel Outcomes in Teaching) specifically aimed at anti-bullying education for the 8–12 age group. The structure of FearNot! was inspired by the Forum Theatre approach developed by Brazilian dramatist Augusto Boal (1979) in order to incorporate theatre into the development of political activism. In this dramatic form, an audience is split into groups, with each group taking responsibility for one of the characters in the drama. Between episodes of dramatic enactment, each group meets the actor, who stays in role, and negotiates with them what they should do next in the drama,

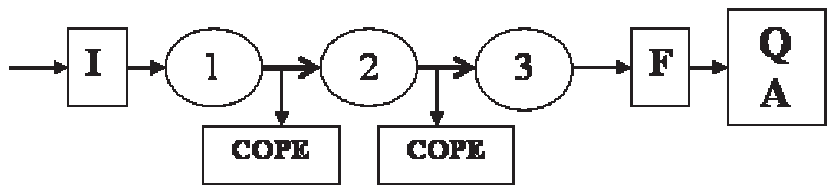


Figure 1. Structure of the FearNot! application



Figure 2. A screen shot from a bullying episode in FearNot!

respecting the constraints of their role and character. This structure of dramatic episodes divided by periods in which advice can be given to a character has been adopted for FearNot! as shown schematically in Figure 1. Here, an introductory scripted segment introduces the characters and school to the child user (I in Figure 1). It is followed by an agent-driven episode in which one of the characters is bullied (a ‘bullying scenario’). At the end of the episode, the victimized character goes to a resources room (the school library) where the child user is asked to give them advice about how to cope with the bullying problem. After a number of episodes (currently three) the drama concludes with an educational message (F in Figure 1) and a questionnaire assessing the extent to which the child user can ‘put themselves in the shoes’ of the characters they have seen and assess their motives and goals (QA in Figure 1). Figure 2 shows a screen shot from one of the bullying episodes in FearNot!.

3. Empathy

The educational objectives of FearNot! depend on the child user being willing to engage with the problems faced by the victimised character. This requires the child to act as an 'invisible friend' – *invisible* because they are not themselves present in the dramatic episodes, and *friend* because they can advise and support the character but not act with god-like power to solve their problems for them. The success of the Japanese Tamagotchis – small plastic capsules with rudimentary graphics expressing their 'needs' for food, cleanliness and play – suggests that children can indeed feel a sense of responsibility for the articulated needs of electronic characters. In the case of FearNot! these needs are emotional rather than physical and for this reason perception of the affective state of the character is fundamental.

In FearNot!, just as with the Tamagotchis, time moves only forwards, and dramatic episodes cannot be replayed, but the repetitive nature of bullying means that at an abstract level all episodes are the same. The psychological basis of the relationship between child and character lies in empathy, and the characters can be described as empathic characters if they succeed in building the desired empathic relationship.

Empathy was a concept originally developed as part of a theory of aesthetics to describe the emotional impact of works of art on the human perceiver. The word is a translation by Titchener (1909) of the term *Einfühlung* of Lipps (1903), which he described as the act of projecting oneself into the object of a perception, 'feeling into' a work of art. A modern definition suggests that empathy is:

any process where the attended perception of the object's state generates a state in the subject that is more applicable to the object's state or situation than to the subject's own prior state or situation. (Preston & de Waal 2002)

One can distinguish between three types of empathy: *cognitive empathy*, *affective empathy* and *ideomotoric empathy*. In the first of these, perception of the 'object' (another person in this case) produces an understanding of their affective state. In the second case, a change in the affective state of the subject is produced: either congruent empathy, in which the resulting state is similar to that of the object, or contrast empathy, in which it is markedly different. Affective empathy in its congruent variant was earlier known as sympathy, but this term has been more recently reserved for the emotional state called compassion (Wispe 1987). Ideomotoric empathy relates to an empathic motor response, and to research that shows that the motoric preactivation of the subject changes due to the perception of the movements of the object (e.g. Prinz 1997).

Apart from different types of empathy, one can also distinguish different mediating mechanisms (Bischof-Köhler 1989). Empathy may be mediated by the situation in which the object is perceived to be, so that for example seeing someone have their handbag stolen may produce the cognitive empathy effect of understanding that they are sad and angry. It may more obviously be mediated by expression, where any element of the full range of expressive behaviour produces the empathic effect. Thus sadness is indicated by crying.

Both of these mediating mechanisms are used in FearNot! However, as in most drama, there is a sense in which situation is dominant in FearNot! and expression is subordinate to it. Thus its focus is less on the accuracy of the expressive behaviour of the characters – as we will see our system is not intended to be naturalistic – and more on using expressive behaviour to emphasise the dramatic situations being presented. In this respect FearNot! is rather different from some of the other applications discussed in this volume, such as online chat or arranging meetings, where expressive behaviour dramatises what is not in itself a very dramatic activity.

4. Expressive behaviour and creating empathy

Much work in expressive behaviour for synthetic characters has focused on producing a naturalistic effect, allied to the parallel move in graphics for photo-realism. Thus the work of Ekman and his colleagues on the coding of human facial expressions (Ekman & Friesen 1978; Ekman 1982), which itself can be traced back to the 19thC photographic work of Duchenne de Boulogne (Duchene 1876), has been a major influence. It has directly impacted the design of the Facial Activation Parameters approach used in mpeg4 and has then been applied to some of the best current work in ‘talking heads’ such as Greta (Martin *et al.* 2006; de Rosis *et al.* 2003), where only the head and shoulders of an agent are shown.

However, in the context of virtual drama, one must question whether believability is the same thing as naturalism. Taking film animation as a guide rather than psychology, of course the answer is a clear negative: good animated characters (Mickey Mouse, Buzz LightYear, Shrek) are highly believable and give ‘the illusion of life’ (Johnston & Thomas 1995) without in any sense being naturalistic. In the same way, theatre uses many non-naturalistic devices, as was demonstrated in a Theatre-in-Education performance on the theme of bullying attended as part of the background research for FearNot! Here a set of short episodes used elements of mime, music, and exaggerated physical action to allow three adult actors dressed in black to portray the life of a victim in a large secondary school.

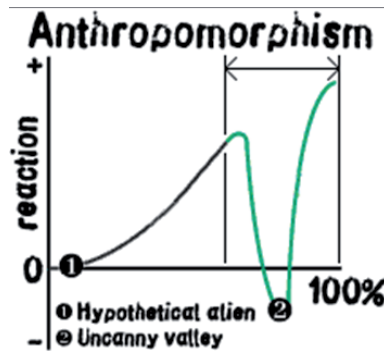


Figure 3. The uncanny valley (after www.arclight.net/~pdb/glimpses/valley.html)

There is also a major pitfall in attempts to produce naturalistic behaviour in autonomous rather than pre-rendered environments not always understood in the graphics community. This was explained by Mori (1982; discussed in Dautenhahn 2002) who examined human reactions to synthetic characters. He argued that the acceptability of such characters increased as they became more anthropomorphic up to a point of near-realism, and then dropped very sharply, right into the negative part of the graph – as shown in Figure 3. He called this drop in acceptability ‘the uncanny valley’ since it appears to correspond to the emotional reaction accompanying shuddering.

It seems that humans develop expectations for synthetic characters based on consistency between appearance and behaviour. The more naturalistic the appearance, the higher the expectations that the behaviour will be naturalistic. Any even slight mismatch between the two is then the cause of the ‘uncanny valley’. Given the subtleties of human expressive behaviour, as for example discussed by LaFrance in this volume, it is clear that meeting the behavioural expectations generated by a very naturalistic appearance is actually very difficult and may be better avoided. Two robot examples illustrate this very well. Kismet (Brezeal 2000) has a metallic face with large red rubber lips, big dark eyebrows and long pink ears a little like those of a donkey and looks nothing like a human baby. Yet analysis of human interaction with it shows that it evokes many of the same behaviours as a baby. On the other hand, the highly anthropomorphic robots of Japanese researchers (Hara & Kobayashi 1996; Mitsunaga *et al.* 2006), with wonderfully engineered and accurate representations of facial muscles, latex skin, glossy hair and glass eyes, can irresistibly remind the observer of zombies.

Both the dramatic paradigm being applied and the dangers of the ‘uncanny valley’ led to particular decisions about the appearance and expressive behaviour to be used in FearNot! Firstly, as can be seen in Figure 2 above and Figure 4, it was

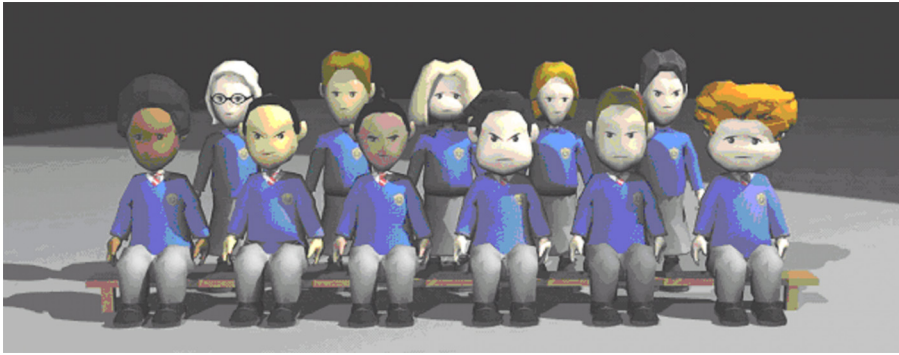


Figure 4. FearNot! cartoon-like characters

decided to make characters cartoon-like in appearance. This reduces expectations about the naturalism of their behaviour and allows non-naturalistic expressive behaviour to be used if needed – in cartoons for example, anger may be signalled by lightning above the head, and crying may create a pool of tears. In fact more conventional combinations of facial expression, stance and animation have been used, but the important point is that less naturalistic behaviour is consistent with the less naturalistic appearance of these characters.

A very simple approach was taken to facial expression in the FearNot! characters. Rather than creating nodes on the facial mesh and animating these to create expressions, a set of facial textures, each displaying a different stereotyped expression were designed, as shown in Figure 5.

The appropriate texture is then displayed on the face when the internal emotional system of the character puts it into the relevant affective state. This may seem unduly simple, but in fact it is very effective – it reduces the need for a child user to decode sophisticated expressions on what are very small graphical faces on a typical desktop computer, and makes it very clear what the emotional reaction of the character is to the situation in which it acts. Added to this is morphing from one expression to the next rather than making an abrupt transition and adding in limited mouth movement when a character speaks (accurate lip synchronisation is another behaviour nobody expects in cartoon-like characters).

Posture is another way in which affective state can be expressed, and its use in FearNot! is illustrated in Figure 6. Some of the ideas expressed in the chapter by Vala and colleagues in this volume, are used here to distinguish between a confident and happy (or angry) character – the bully – and a demoralised and sad character – the victim. In this cartoon-like milieu, exaggeration is entirely possible – as we have seen with the facial expressions – and both exaggerated postures (as part of animations) and exaggeration via the use of shadow have been used in FearNot!

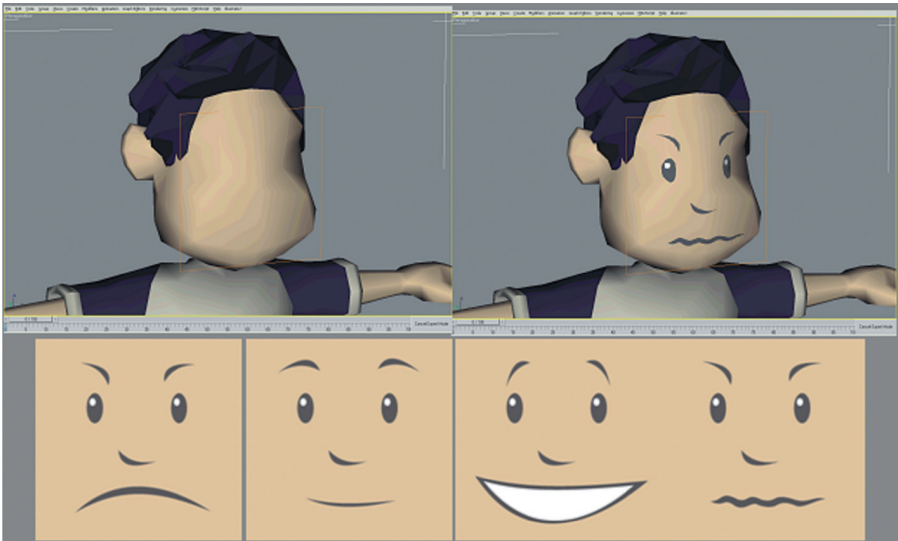


Figure 5. Producing facial expressions in texture



Figure 6. Using expressive posture. Which one is the bully?

FearNot! seeks to produce unscripted dramas, in which action is generated by interaction among the characters. In a scripted system, the required expressive behaviour can be invoked at the global level and, for example, transmitted to characters via a mark-up language such as Behaviour Markup Language (BML – Kopp *et al.* 2006) or those cited in the chapter by Ruttkay and colleagues in this volume. In an unscripted system, expressive behaviour must be generated, like all other actions, by the internal architecture of each character through a sense-reflect-act cycle, so that characters produce their own mark-up in real time. This is

Table 1. Associating actions with OCC emotions in the bullying case

OCC emotion	Appropriate actions/behaviours
Joy	Smile, dance, laugh, wave
Happy-for	Felicitate, encourage
Sorry-for	Apologise, encourage, protect
Anger	Ignore, hit, avoid, aggress, humiliate
Distress	Cry, sit on the floor, beg

feasible, though extremely difficult, for facial expression and posture, but the use of dramatic lighting, shadows, close-ups, cutting between characters and other dramatic effects depends on the development of intelligent camera agents, outside the scope of this project.

We will not discuss the character action selection system in detail here (see Aylett *et al.* 2006), but in summary, it uses appraisal rules as in the well-established taxonomy of Ortony, Clore and Collins (1988) to establish the emotional reaction of a character to an event (an action of another character) or object. This is then used both to generate reactions – such as expressive behaviour and impulsive physical actions – and to establish goals for which immediate action sequences are planned. A memory of recent interactions and a parameterised set of personality parameters will also impact action selection. Table 1 shows some of the actions or behaviours associated with relevant OCC emotions in the bullying scenarios. The personalities of the characters are parameterised according to which of five roles they play: bully, victim, bully-victim, defender or bystander. Figure 7 below shows an extract from the configuration file for a bully, expressed in XML.

The extract shows values assigned to *emotional disposition*, *goals*, and *event response*. The *EmotionDisposition* tag is used to specify the character's emotional threshold (0 = low resistance to the emotion and 9 = high resistance) and decay rate (0 = long duration of emotional state, 9 = short duration) for each of twenty-two emotion types, a subset of those defined in the OCC system seen as relevant for bullying scenarios. In the Figure 7, for example, the bully does not easily feel afraid since a high threshold value (9) makes the character less sensitive to the emotion. Even if a fear emotion is created, it will disappear quickly, since the character has a very high decay rate (8). The *Goal* tag specifies what the character's goals are and how important they are to him. The goal described is the goal of physically bullying the victim. This importance of success in this goal is very high to the bully, and if he fails to achieve the goal, he will also feel very troubled. Since a goal can be shared by more than one character (but with different degrees of importance), a file containing all goals is used. In this way, each goal is defined only once. Finally, the *Event* tag defines an emotional reaction rule, used in respond-

```

<Character id='1' role='Bully'>
  <EmotionDisposition emotion='Fear' threshold='9' decay='8' />
  ...
  <Goals>
    ...
    <Goal id='4' name='BullyVictim'
      importanceOfSuccess='9' importanceOfFailure='9' />
    ...
  </Goals>
  <EventReactions>
    ...
    <Event type='Action' subject='OTHER' action='Cry'>
      <Desirability value='9' />
      <DesirabilityForOther value='-10' />
      <Praiseworthiness value='-5' />
    </Event>
    ...
  </EventReactions>

```

Figure 7. Extract from the role-definition of a bully

ing to emotional states of other characters. The example in Figure 7 specifies the bully's emotional reaction when he sees another character crying. He will desire such an event, but knows that it is very undesirable to the character that performs the action. Also, according to his standards, men do not cry and someone who cries is no more than a wimp. Therefore he sees the cry action as blameworthy.

It is worth noting here that to some extent we disagree with Carofiglio and colleagues (this volume) when they say: "If, for instance, the application concerns a 2D embodied character that is sketchy in its appearance and is expected to show a limited range of expressions, a refined modelling approach is probably not needed." The FearNot! Characters, though not 2D, are indeed very sketchy in appearance, but in order to create dramatic interaction a heavy onus is placed on their individual action selection systems, requiring a much more sophisticated emotional model than is visible in their expressive behaviour. The important point here is that in an emotionally-driven drama, all action is to some extent expressive of the inner state of the characters because it is interpreted through the empathic relationship between the child user and the victim character, mediated by the dramatic situation. Thus the link between internal character state and selected action must be appropriate in order to produce coherent interaction between the characters.

Table 2. Results of preliminary evaluation with 127 children

Movement	1= +ive view 5= -ve view	Emotion towards characters	Luke (bully)	John (victim)	Martinha (narrator)
believable	3.04	Sadness – % feeling	5	95	0
realistic	3.11	Anger – % feeling	85	13	2
smooth	2.82				
Attractiveness of school environment	2.1	Match with characters	2.0		

5. Evaluation of FearNot!

A large-scale evaluation of FearNot! took place in the summer of 2004 with about 400 children aged 9–11 in the UK. The individual responses as expressed through questionnaires are discussed elsewhere (Hall *et al.* 2006), but the focus groups held at the end of each of seven session revealed a very positive response from the child-users to using FearNot! This appears to support earlier evaluation results that were obtained by using what was known as the ‘trailer’ for FearNot! – a small-scale precursor to the overall application consisting only of the introductory segment and a single bullying episode, with no interaction with the victim character.

While the results of the trailer evaluation are also discussed in detail elsewhere (Hall *et al.* 2004), we here examine (Table 2) the responses to the appearance and movement of the characters and the amount of empathy felt by subjects with the victim. The questionnaire focused on character attributes (voice believability, likeableness, conversation content, movement) storyline (believability), character preferences and empathy (sorrow and anger). Measurement was predominantly by a 5 point Likert scale and the questionnaire was completed by 127 children from schools in England and Portugal, 64 male and 63 female, aged from 8–13 ($x = 9.83$, $SD = 1.04$). They were drawn from primary schools located in urban and rural areas of Hertfordshire, UK (47%) and Cascais, Portugal (53%).

The results show that for example the children were not specifically impressed by the movement of the characters: 1 was a positive response and 5 a negative, and the values for believability, realism and smoothness were all around 3. However, the 3D school environment and its match to the style of the characters was valued more highly, with results around 2. The questions on empathy indicate clearly that children did empathise with the victim, with very high percentages feeling sorry for the victim and angry with the bully. We take this as indicating that the design decisions on expressive behaviour, when taken with the narrative action, have supported the empathic engagement we were seeking.

Issues of culture and gender have also been highlighted in the evaluation work carried out so far. Use of story-board prototypes has indicated that while girls empathise with characters of both genders, boys empathise with male characters. A complicating factor here is that the scenarios involving girls tended to involve relational bullying rather than straightforward physical bullying, and questions on story-content indicated that boys in this age group found it very difficult to understand relational bullying or indeed to see it as bullying at all.

6. Conclusions

As already commented, FearNot! is rather different in its approach from the embodied conversational characters (ECAs) discussed elsewhere in this volume. The synthetic characters involved are essentially virtual actors, and drama rather than naturalism is therefore the key paradigm. However there are also key differences between FearNot! and two other applications which are much closer to its objectives.

The Mission Rehearsal Exercise (MRE) (Swartout *et al.* 2001) has a similar dramatic element in that it uses a traffic accident during a peace-keeping exercise as a way of producing training situations for soldiers. However, unlike FearNot!, it is intended as an immersive experience and rather than using an empathic relationship as a pedagogical approach it creates the sort of emotional stress that a soldier who has conflicting objectives and little time to resolve them might experience. The user is a trainee who must make decisions in real time and therefore has a higher cognitive load than that imposed on FearNot!'s child users during its dramatic episodes. Additionally, as a military training application, it is required to move a great deal closer to realistic representation than is needed in an application for children. Much military training is intended to reduce the amount of expressive behaviour produced in stressful situations, so that its military characters remain fairly deadpan. The key expressive figure in the scenario is that of the mother of the injured child, and here posture, gesture and pre-recorded voice are used to produce dramatic expressiveness and to increase the stress upon the trainee.

A second relevant application is Carmen's Bright IDEAs (Gratch & Marsella 2001), which is aimed at mothers of children with cancer. Its interactive component shows a dialogue between a mother called Carmen and a therapist called Gina, and the pedagogical objective is to teach a real-life mother a cognitive problem-solving technique. In the same way as FearNot!, it tries to meet its pedagogical objective by creating empathy between the user and Carmen, and allows the course of the interaction to be changed by putting up three 'thought bubbles' for Carmen at various points from which the user must select one. As in FearNot!,

a cartoon-like style has been adopted, though this is implemented in 2D rather than 3D. However, a more significant difference is that the episodes are cast as conversations with a therapist, so that the dramatic dynamic (as well as the action selection issue) is rather different. The amount of facial expressiveness is if anything more limited than in *FearNot!*, but pre-recorded voices contributed by actors make the aural channel extremely important as an expressive medium.

We have said little about the use of expressive voice in *FearNot!* so far, and it is worth pointing out that it is a modality absent also from other discussions in this volume. The reason is that a number of problems remain to be solved before synthesised speech can be used successfully for expressive purposes. Artificially-synthesised voices still sound very ‘robotic’ to the point that their use in affectively expressive situations is likely to be counter-productive. A much more successful approach is based on unit-selection, in which speech is constructed by concatenating segments from pre-recorded human input. These are much more human-sounding and believable, but having been aimed at applications such as automatic telephone systems, tend to be equable rather than emotional in tone. In addition, they require a substantial memory allocation to hold a large real-time database making them problematic alongside a 3D real-time rendering system. *FearNot!*, like the two other applications discussed in this section, used pre-recorded speech in its trailer; however, this approach is not compatible with the unscripted dramas for which *FearNot!* is aiming. In its most recent version, *FearNot!* has incorporated unit-selection based speech (Weiss *et al.* 2007), but this is still to be evaluated.

To conclude, we have tried to demonstrate in this chapter that consistency may be a more important objective in expressive behaviour for synthetic characters than outright realism. Believability – the willingness of a user to accept a collection of graphical pixels as a personality with ‘its own’ internal state, is the key we would argue to successful user-agent interaction. In *FearNot!*, we believe we have created characters with which child users do engage and which can indeed produce the empathic engagement needed to meet its pedagogical objectives.

Acknowledgements

The work on *FearNot!* discussed was partially supported by European Commission (EC) under the Information Society Technologies (IST) RTD programme in the project VICTEC contract IST-2001-33310-victec (www.victec.net) and is currently funded by the eCIRCUS project IST-4-027656-STP (www.e-circus.org). The authors are solely responsible for the content of this publication. It does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of data appearing therein.

References

- Argyle, M. (1972). Non-verbal communication in human social interaction. In R. A. Hinde (Ed.), *Non-verbal communication* (pp. 243–269). Cambridge: Cambridge University Press.
- Aylett, R. S., Dias, J. & Paiva, A. (2006). An affectively-driven planner for synthetic characters. In *Proceedings of The International Conference on Automated Planning and Scheduling, ICAPS 2006* (pp. 2–10), June 6–10, 2006, The English Lake District, Cumbria, UK. AAAI Press.
- Bates, J. (1994). The Role of Emotion in Believable Agents, *Communications of the ACM* 37(7), 122–125.
- Boal, A. (1979). *The Theatre of the Oppressed*. New York: Urizen Books.
- Bischof-Köhler, D. (1989). *Spiegelbild und Empathie*. Bern: Huber.
- Breazeal, C. (2000). Infant-like Social Interactions between a Robot and a Human Caregiver. *Adaptive Behavior*, 8, 9–75.
- Crick, N. R. & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development*, 66, 710–722.
- Dautenhahn, K. (2002). The Design Space of Life-Like Robots. In D. Polani, J. Kim, & T. Martinetz (Eds.), In *Proc. 5th German Workshop on Artificial Life* (pp. 135–142). Amsterdam: IOS Press.
- De Rosis, F., Pelachaud, C., Poggi, I., De Carolis, N. & Carofiglio, V. (2003). From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59(1–2), 81–118.
- Duchenne (de Boulogne), G.-B. (1876). *Mecanisme de la Physionomie Humaine: Atlas*. Deuxieme edition. Paris: J.-B. Bailliere et Fils.
- Ekman, P. & Friesen, W. V. (1978). *Manual for the Facial Action Coding System*. Palo Alto, CA: Consulting Psychology Press.
- Ekman, P. (1982). *Emotions on the Human Face*. Cambridge University Press.
- Gratch, J. & Marsella, S. (2001). Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents. In J. P. Muller, E. André, S. Sen & C. Frasson, *Proc. Fifth Intl. Conference on Autonomous Agents* (pp. 278–285). New York, NY: ACM Press.
- Hara, F. & Kobayashi, H. (1996). A Face Robot Able to Recognize and Produce Facial Expression. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS '96* (pp. 1600–1607). Osaka, Japan, 4–8 November, 1996.
- Hall, L., Woods, S., Dautenhahn, K., Sobral, D., Paiva, A., Wolke, D. & Newall, L. (2004). Designing empathic agents: Adults vs Kids. In James C. Lester, Rosa Maria Vicari & Fábio Paraguaçu (Eds.), *Intelligent Tutoring Systems, 7th International Conference, ITS 2004* (pp. 604–613) [LNCS 3220]. Heidelberg & Berlin: Springer-Verlag.
- Hall, L., Woods, S. & Aylett, R. S. (2006) Using Theory of Mind Methods to Investigate Empathic Engagement with Synthetic Characters. *International Journal of Humanoid Robotics: Special issue Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids*. Vol. 3, No. 3 (2006), 351–370.
- Johnston, O. & Thomas, F. (1995). *The illusion of life: Disney animation*. New York: Hyperion Press.
- Kopp S., Krenn B., Marsella S., Marshall A., Pelachaud C., Pirker H., Thorisson K. & Villhjalmsjon, H. Towards a Common Framework for Multimodal Generation: The Behavior

- Markup Language. In J. Gratch *et al.* (Eds.), *Intelligent Virtual Agents 2006* [LNAI 4133] (pp. 205–217). Heidelberg & Berlin: Springer-Verlag.
- Lipps, T. (1903). *Ästhetik*. Teil 1. Hamburg, Leipzig.
- Martin, J.-C., Niewiadomski, R., Devillers, L., Buisine, S. & Pelachaud, C. (2006). Multimodal complex emotions: Gesture expressivity and blended facial expressions, *International Journal of Humanoid Robotics*, 3(3), 269–291.
- Mitsunaga, N., Miyashita, T., Ishiguro, H., Kogure, K. & Hagita, N. (2006) Robovie-IV: A Communication Robot Interacting with People Daily in an Office. *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'06)*, CD-ROM, Beijing, China, Oct. 2006.
- Mori, M. (1982). *The Buddha in the Robot*. Boston, Tokyo & Singapore: Charles E. Tuttle Co.
- Olweus, D. (1978). *Aggression in Schools: Bullies and Whipping Boys*. Washington, DC: Hemisphere.
- Olweus, D. (1991). Bully/victim problems among schoolchildren: Basic facts and effects of a school-based intervention programme. In D. Pepler & K. Rubin (Eds.), *The Development and Treatment of Childhood Aggression* (pp. 411–438). Hillsdale, NJ: Erlbaum.
- Ortony, A., Clore, G. L. & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.
- Preston, S. D. & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1–71.
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9, 129–154.
- Swartout, W., Hill, R. W. Jr., Gratch, J., Johnson, L. W., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J. F., Rickel, J., Thiébaux, M., Tuch, L. & Whitney, R. (2001). Toward the Holodeck: Integrating Graphics, Sound, Character and Story. In J. P. Muller, E. Andre, S. Sen & C. Frasson (Eds.), *Proceedings of 5th International Conference on Autonomous Agents* (pp. 409–416). New York: NY: ACM Press.
- Titchener, E. (1909). *Experimental psychology of the thought processes*. New York: Macmillan.
- Weiss, C., Oliveira Luis C., Vogt Th., Andre, E., Vala M., Paiva, A. & Hall, L. (2007). eCircus: Spoken Interaction of Autonomous Agents in Educational Virtual Environments. In *Proc. 33rd German Annual Conference on Acoustics, DAGA 2007*. 19–22 March, 2007, Stuttgart, Germany.
- Wispe, L. (1987). History of the Concept of Empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and Its Development* (pp. 17–37). Cambridge: Cambridge University Press.
- Wolke, D., Woods, S., Schulz, H. & Stanford, K. (2001). Bullying and victimisation of primary school children in South England and South Germany: Prevalence and school factors. *British Journal of Psychology*, (92), 673–696.

Psychological and social effects to elderly people by robot-assisted activity

Takanori Shibata, Kazuyoshi Wada, Tomoko Saito
and Kazuo Tanie

1. Introduction

1.1 Aged society

According to the United Nations, if the proportion of people 65 years old and over in the population of a country exceeds 7%, this indicates an aging society, and if that proportion exceeds 14%, an aged society. Figure 1 shows the changing proportions in most advanced countries. Countries other than the US have become aged societies (United Nations 1998). Such percentages are expected to increase, together with the number of elderly people who require nursing due to dementia, becoming bedridden, etc., as well as the number being institutionalized for long periods in care facilities for the elderly. Moreover, the bodily and mental stress of nursing staff, occasioned by manpower shortages and increasing workloads, is becoming a big problem. Mental stress in nursing causes Burnout syndrome (Maslach 1976) and makes nursing staff irritable, with loss of sympathy for their patients. Thus, it is important to improve the “quality of life (QOL)” of elderly people, as this helps them to live healthy and independent lives. It also saves the social costs of supporting elderly people.

1.2 Animal-assisted therapy and activity

Interaction with animals has long been known to be emotionally beneficial to people. In recent years, the effects of animals on humans have been researched and proved scientifically. Friedmann investigated the one-year survival of patients who were discharged from a coronary care unit, finding that survival amongst those who kept pets was higher than those who did not (Friedmann *et al.* 1980). Baum *et al.* reported that blood pressure dropped when people petted their dog (Baun *et al.* 1984). Garrity and colleagues (1989) studied elderly people who were

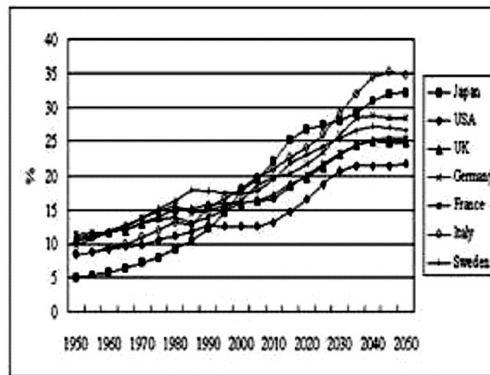


Figure 1. Ratio of people 65 years old and over against total population of most advanced countries

socially isolated and lost their partner within the previous year, and found that the depth of depression among those who had no pets was higher than those who did. Lago and colleagues (1989) investigated the influences of pet owning on elderly people through telephone interviews. They discovered that mortality and attrition were higher for former owners than current owners. Hart and colleagues (1987) studied the social influences of animals on people. They found that the number of friendly approaches by strangers to people with dogs were greater than to people without the dogs.

In medical applications, especially in the United States, animal-assisted therapy and activities (AAT&AAA) are becoming widely used in hospitals and nursing homes. AAT has clear goals set out in therapy programs designed by doctors, nurses or social workers, in cooperation with volunteers. In contrast, AAA refers to patients interacting with animals without particular therapeutic goals, and depends on volunteers. AAT and AAA are expected to have 3 effects:

1. Psychological effect, e.g. relaxation, motivation;
2. Physiological effect, e.g. improvement of vital signs;
3. Social effect, e.g. stimulation of communication among inpatients and caregivers.

For example, a hospitalized child, who was in significant pain because of his disease, was afraid to get up and walk around. However, when he was asked to take a therapy dog for a walk, he immediately agreed and walked off happily, as if his pain had diminished. Moreover, the dog acted as a medium for interaction between him and the other children (Kale 1992). In another case, a boy who was born to a drug-using mother, as a crack-exposed baby could not speak and walk.

However, through interaction with therapy dogs and birds, he improved both his linguistic and motor ability (Delta Society 1991).

In the case of people with AIDS, it is particularly important to reduce their stress because there is a strong relationship between the complications of immune deficiency and stress. AAT brings the effects of relaxation to them, and helps them to stay connected with the world (Haladay 1989).

In addition to these effects, AAT and AAA at nursing homes provides the effects of rehabilitation to elderly people, and offers laughter and enjoyment to a patient who has little remaining life (Gammonley & Yates 1991). Moreover, there are some cases where the therapy has improved the state of elderly people with dementia.

However, most hospitals and nursing homes, especially in Japan, do not accept animals, even though they admit the positive effects of AAT and AAA. They are afraid of the negative effects of animals on human beings, such as allergy, infection, bites, and scratches.

1.3 Mental commitment robot

Science and technology has been developed on the basis of objectivity, and this development has been designed for universality and commonality. "Technology," as an application of the practice of science, is the skill of modifying and processing events in nature so that they become useful to human life. Science, which is the foundation for technology, is a body of knowledge which is systematic and empirically verifiable. On the other hand, "art" is the activity and ability of humans who attempt to create and express aesthetic values by making full use of certain materials, techniques, and methods. It is industrial arts that creates practical objects with aesthetic value.

Robotics has been applied to automation in industrial manufacturing. Most robots are machines for optimizing a practical system in terms of objective measures such as accuracy, speed and cost (Petroski 1996). Therefore, humans give machines suitable methods, purposes and goals (Sánchez *et al.* 1997; Shibata *et al.* 1996a). Machines are the passive tools of humans.

We have been researching robots that contrast with such machines. If a robot were able to generate its own motivations and behave voluntarily, it would have significant influence over an interacting human. At the same time, the robot would not be a simple tool for the human nor could it be evaluated only in terms of objective measures. Figure 2 depicts the space of objective and subjective evaluations classified by application in the design of artificial objects.

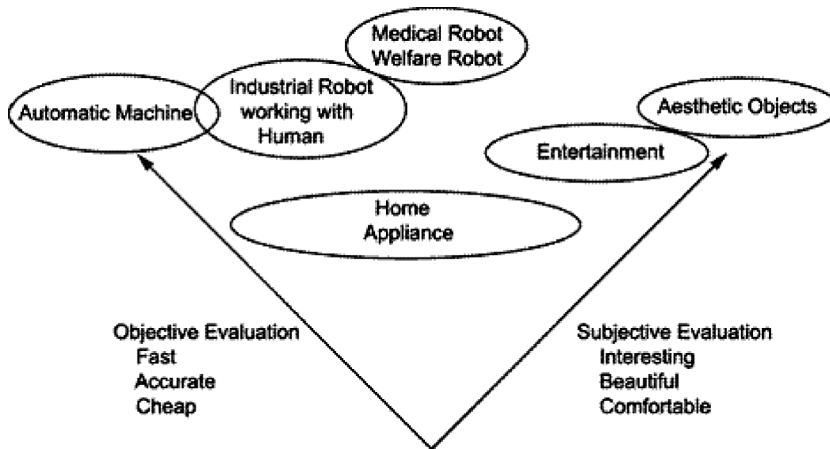


Figure 2. Objective and subjective evaluations classified by application in the design of artificial objects

We have been designing animal type robots as examples of artificial emotional creatures (Shibata *et al.* 1996b, 1997, 1999, 2000). These animal-type robots have physical bodies and exhibit active behavior whilst generating goals and motivations by themselves. They interact with human beings physically. People recognize the robots and subjectively interpret their movements based on their knowledge and experience. When we engage physically with an animal type robot, it stimulates our affection. We then experience positive emotions such as happiness and love, or negative emotions such as anger and fear. Through physical interaction, we develop an attachment to the animal type robot, while regarding it as either intelligent or stupid from our subjective perspective. In this research, animal type robots that give mental value to human beings are referred to as “mental commitment robots.” Three examples that we have developed are dog, cat and seal robots.

1.4 Robot-assisted therapy and activity

We have proposed robot therapy (Shibata *et al.* 1996b) termed robot-assisted therapy (RAT). We have used a seal robot named Paro (Figure 3), instead of real animals, in pediatric therapy at a university hospital (Shibata *et al.* 2000).

The children’s ages were from 2 to 15 years, some of them having immunity problems. Paro was given to them three times a day for 11 days. The children’s moods improved on interaction with Paro, encouraging the children to communicate with each other and with carers (Figure 4, left). In one striking instance, a

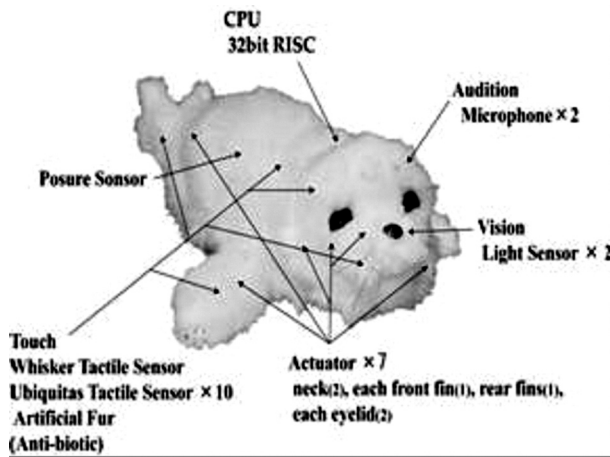


Figure 3. Seal-robot “Paro”



Figure 4. Interaction between inpatient and Paro (left), and with elderly people (right)

young autistic patient recovered his appetite and his speech abilities during the weeks when Paro was at the hospital. In another case, a long-term in-patient felt pain when she moved her body, arms, and legs, and could not move from her bed. However, when Paro was given to her, she smiled and was willing to stroke Paro. A nurse said that Paro had a rehabilitative function as well as a mental effect.

In addition, seal robots have been applied to robot-assisted activity (RAA) for elderly people in a day service center (Shibata *et al.* 2001a; Wada *et al.* 2004; Saito *et al.* 2002a). The day service center is an institution that aims to decrease the nursing load on a family by caring for elderly people during the day. Interaction with the seal robots improved their moods, encouraging them to communicate with each other and carers (Figure 4, right). Moreover, the results of urinary tests

indicated that interaction with Paro reduced stress in the elderly. In an interesting instance, an elderly person who had seldom talked with other people, started to talk volubly with others when she was interacting with the seal robot. In addition, the seal robot had influences on people who suffered from dementia. One example is that of an elderly person who did not attempt to behave independently, and often forgot things that she had just done. When she was interacting with the seal robot, she often laughed and seemed to be brighter than usual. Another example is that of an elderly person who tended to want to go back home, but she remained at the day service center to play with the seal robot, and looked happy. Also investigated was the mental load on nursing staff when nursing elderly people. The results showed that their mental load decreased because the elderly people would spend their time by themselves with the seal robots.

We have investigated long-term interactions between Paro and the elderly, and found that the effects of interaction with Paro lasted for more than one year (Wada *et al.* 2005a). Furthermore, the neuropsychological effects of Paro on patients with dementia were assessed by analyzing their EEGs (Wada *et al.* 2005b). The results showed that interaction with Paro improved the activity of the patients' cortical neurons, especially for those who liked Paro. Studies have also been conducted using questionnaires given out at exhibitions held in six countries (Japan, U.K., Sweden, Italy, Korea, and Brunei), to investigate how people evaluate the robot. The results showed that the robot was widely accepted, regardless of cultural differences (Shibata 2004).

Other animal-type robots such as Furby, AIBO (Fujita & Kitano 1998), NeCoRo, etc., have been released by several companies. Robot-assisted activity that uses those robots has also been tried. For example, Yokoyama used AIBO in a pediatrics ward, and observed the interaction between children and the AIBO (Yokoyama 2002). He pointed out that the stimulus received from AIBO was strong; however, the stability was quite weak, compared with living animals. In other words, when people meet AIBO for the first time, they are interested in it for a while. However, relaxation effects such as those obtained from petting a real dog were never felt from AIBO.

This chapter addresses the application of seal robots to assist the activity of elderly people at a health service facility for the aged, in order to investigate the psychological and social effects of seal robots on those elderly people who stayed at the facility. The effects of the seal robot were then compared with those of a placebo seal robot having a modified motion generation program. Section 2 explains how a seal robot and a placebo seal robot were used for RAA. Section 3 describes the uses of RAA for elderly people. Section 4 explains the effects of RAA. Section 5 discusses current results of RAA and future work. Finally, Section 6 offers conclusions.

2. Seal robot and placebo seal robot

2.1 Specifications of our seal robot

Our seal robot, Paro, was developed to physically interact with human beings (Figure 3). Paro's appearance is that of a baby harp seal, which is a non-familiar animal for our users. Therefore, people can accept Paro easily without preconceptions. As for perception, Paro has tactile, vision, auditory, and posture sensors beneath its soft white artificial fur. In order that Paro should have a soft body, an air-bag type tactile sensor was developed and implemented. To provide movement, the robot has seven actuators; two for eyelids, two for the neck, one for each front fin, and one for two rear fins. Paro weighs about 2.8 kilograms. It has a behavior generation system consisting of two hierarchical layers of processes: proactive and reactive. These two layers generate three types of behavior; proactive, reactive, and physiological behaviors. Figure 5 depicts Paro's behavior generation system.

Proactive behavior

Paro has two layers to generate its proactive behavior: a behavior-planning layer and a behavior-generation layer. By addressing its internal states of stimuli, desires, and a rhythm, Paro generates proactive behavior.

Behavior planning layer. This contains a state transition network based on the internal states of Paro and Paro's desires, produced by its internal rhythm. Paro

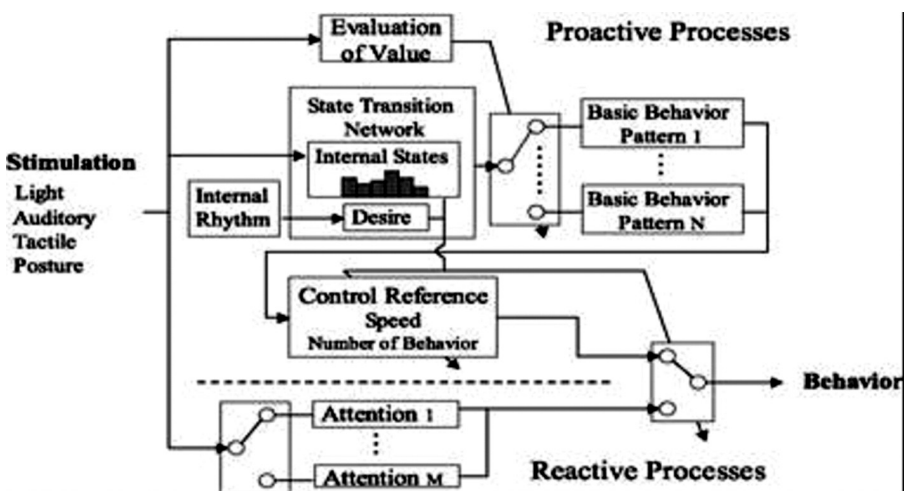


Figure 5. The behavior generation system of Paro

has internal states that can be named with words indicating emotions. Each state has a numerical level which is changed by stimulation. The state also decays over time. Interaction changes the internal states and creates the character of Paro. The behavior-planning layer sends basic behavioral patterns to the behavior-generation layer. The basic behavioral patterns include several poses and movements. Here, although the term “proactive” is used, the proactive behavior is very primitive compared with that of human beings. The behavior of a real seal was observed from commercial videos. Additionally, the habitat of real seals was visited to investigate their ecology. Then, the behavior in Paro was implemented to be similar to that of a real seal.

Behavior generation layer. This layer generates control references for each actuator to perform the determined behavior. The control reference depends on the magnitude of the internal states and their variation. For example, parameters can change the speed of movement, and the number of instances of the same behavior. Therefore, although the number of basic patterns is finite, the number of emerging behaviors is infinite because of the varying number of parameters. This creates life-like behavior. In addition, to gain attention, the behavior-generation layer adjusts parameters relating to the priority of reactive behaviors and proactive behaviors based on the magnitude of the internal states. This function contributes to the behavioral situation of Paro, and makes it difficult for a subject to predict Paro’s action.

Long-term memory. Paro incorporates reinforcement learning. It has a positive value for preferred stimulations such as stroking. It also has a negative value for undesired stimulations such as beating. Paro assigns values to the relationship between stimulation and behavior. Gradually, Paro can be tuned to the preferred behaviors of its owner.

Reactive behavior

Paro reacts to sudden stimulation. For example, when it hears a sudden loud sound, Paro pays attention to it and looks in the direction of the sound. There are several patterns of combination of stimulation and reaction. These patterns play the role of conditioned and unconscious behavior.

Physiological behavior

Paro has a diurnal rhythm. It has several spontaneous needs, such as sleep, based on this rhythm.

2.2 Specifications of the placebo seal robot

It is frequently experienced that interest is lost in toys once the mechanics of their operation are understood. Consider, then, the following hypothesis:

People can rapidly predict the action of robots that execute only defined simple motions, with consequent loss of interest, leading to loss of effectiveness of the robot.

Following this hypothesis, the regular Paro program was modified to create a placebo Paro as follows.

Proactive behavior

Repetition of following five types of action:

1. Blinking
2. Swinging the rear fins to right and left
3. Swinging both front fins forwards and backwards
4. Swinging the head to right and left
5. Crying → Return to (1)

Reactive behaviors

The following simple reactions to stimuli:

1. Crying (sound is different from the cry associated with proactive behavior)
2. Raising of the head

3. Robot assisted activity for elderly people

Paro was applied to robot-assisted activity for elderly people at a health service facility for the aged, in order to investigate the effects on elderly people. The health service facility for the aged is an institution that provides several services, such as long stays in the institution, day care and rehabilitation of elderly people. People needing nursing can remain there for a certain period. In order to rehabilitate back into society, they are provided with daily care and trained to spend their daily life independently during their sojourn. At the start of the experiment at the institution, about 100 elderly people were residing there. Moreover, about 30 of these people had dementia. People who did not have dementia stayed in buildings A and B. On the other hand, people who had dementia stayed in building C, and were isolated from other people. Usually, they did not communicate with each other very much; the atmosphere was gloomy.

Before starting the robot-assisted activity, the purposes and means of the experiment were explained to the elderly people who resided in buildings A and B, and received their approval. The elderly people who approved the investigation exhibited a variety of symptoms (no answer to questionnaires, bedridden, etc). Some people were impossible to question. In these cases, nursing staff that were well aware of the usual state of these elderly people evaluated them, and decided who could be investigated. After the evaluation, 23 subjects were chosen. 12 subjects (4 men and 8 women) resided in building A, and 11 subjects (2 men and 9 women) were in building B. Their average age with standard deviations were 84.6 ± 7.0 and 85.5 ± 5.4 for men and women, respectively.

3.1 Activity program

A regular Paro was provided for the subjects who resided in building B, and a placebo Paro was provided for the subjects in building A. In order to prevent the subjects of one group interacting with the Paro of the other group, they interacted with the relevant Paro while segregated in the institution. Moreover, the existence of two kinds of Paro was kept secret from subjects. Each group interacted with each Paro for about one hour at a time, four days a week, for three weeks. A desk was prepared to set Paro in the center of people, and the subjects were arranged as shown in Figure 6. However, all the subjects could not interact with Paro at the same time. Therefore, Paro was moved among subjects in turn, ensuring the same interaction time with Paro for each subject.



Figure 6. Interaction between subjects and Paro

3.2 Methods of evaluation

In order to investigate elderly people's moods before and after introduction Paro to the institution, the following types of data and additional information were collected:

1. Face scale (Lorish & Maisiak 1986) (Figure 7)
2. Profile of Mood States (POMS) (McNair *et al.* 1992)
3. Comments from nursing staff

The Face Scale contains 20 drawings of a single face, arranged in serial order by rows, with each face depicting a slightly different mood state. A graphic artist was consulted so that the faces would be portrayed as genderless and multiethnic. Subtle changes in the eyes, eyebrows, and mouth were used to represent slightly different levels of mood. They are arranged in decreasing order of mood and numbered from 1 to 20, with 1 representing the most positive mood and 20 representing the most negative mood. As the examiner pointed at the faces, the following instructions were given to each patient: "The faces below go from very happy at the top to very sad at the bottom. Check the face which best shows the way you feel inside now."

POMS is one of a set of popular questionnaires to measure a person's mood. POMS are used in various research fields such as medical therapy and psycho-

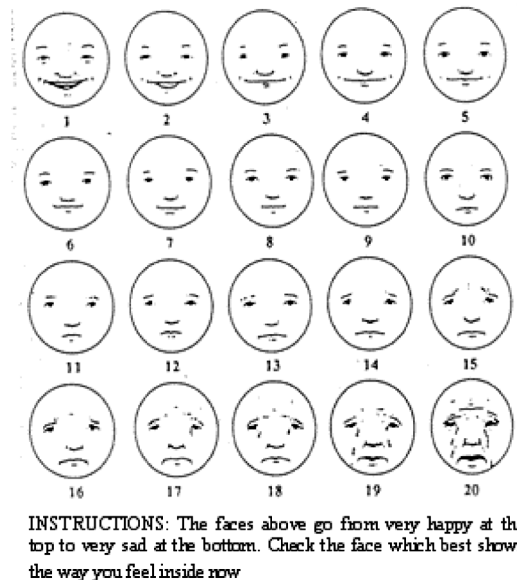


Figure 7. Face Scale

therapy. It can measure six mood states at the same time: Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigor, Fatigue, and Confusion. It contains 65 items concerning moods. Each item is evaluated in five stages from 0–4: 0 = not at all, 1 = a little, 2 = moderately, 3 = quite a bit, and 4 = extremely. Of the 58 items (7 of which are dummy items), 65 items are classified into the six mood states, and the total scores of each mood state are then calculated. The total scores are then translated into standard scores by using a special table. The face scale and POMS were applied to the subjects one week before introduction of Paro, and again the 2nd and 3rd week after introduction. In addition, familiarity with Paro was investigated once a week through questionnaires. The questionnaires comprised 3 items: I like Paro, I speak to Paro, and Paro is like a child or grandchild for me. These items were evaluated over five stages: 0 = not at all, 1 = a little, 2 = moderately, 3 = quite a bit, and 4 = extremely.

4. Results of robot assisted activity

Figure 8 (left) shows the average face value. The average scores of the regular Paro group decreased from about 9.0 (before introduction) to 7.0 (3rd week). In addition, the average scores of the placebo Paro group also decreased from about 7.0 (before introduction) to 6.3 (3rd week). As for POMS, Figure 8 (right) shows average standard scores of Depression-Dejection. Here, 50 standard points represent the average score of Depression-Dejection of Japanese people over 60 years old. Average standard scores of the regular Paro group decreased from about 61 (before introduction) to 47 (3rd week). Moreover, the average standard scores of placebo Paro group also decreased from about 58 (before introduction) to 51 (3rd week).

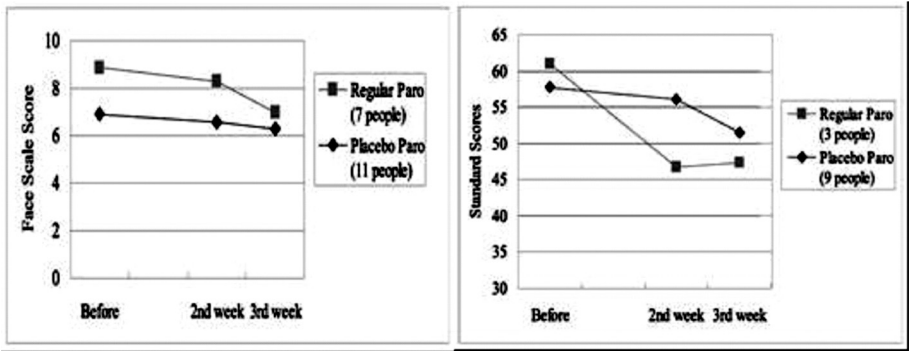


Figure 8. Average face scale Scores (left) and standard scores of “Depression-Dejection” of POMS (right) of elderly people over 4 weeks

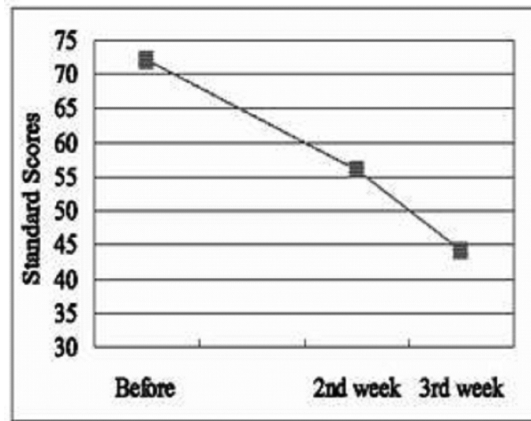


Figure 9. Standard scores of “Depression-Dejection” of POMS of a 96-year-old male over 4 weeks

In one striking instance, Figure 9 shows the result of a subject who was a 96-year-old male. He was usually not sociable, and even carers could hardly communicate with him. Before the introduction of Paro, his standard score of “Depression-Dejection” of POMS was very high. However, after introduction of Paro, he liked Paro very much. He often laughed and sang songs to Paro when he was interacting with it. Then, he made the surrounding people laugh. Carers were surprised by his change. Moreover, his standard scores dramatically decreased to 44 in the 3rd week after the introduction of Paro. According to the comments and observations of nursing staff, both groups of subjects eagerly awaited Paro and willingly participated in interaction with Paro. Paro increased their laughter, and encouraged subjects to communicate both with each other and the nursing staff. In an interesting instance, an elderly woman who liked Paro very much made a song for Paro and sang it to Paro. Then, she looked very happy.

Figure 10 shows results of the average scores of “I like Paro”, being one of the questionnaires’ items of familiarity with Paro. The average score of the regular Paro group decreased from about 3.0 to 1.5. On the other hand, the average score of the placebo Paro group remained at a high value of about 3.0 for three weeks. As a statistical analysis, Friedman’s test was applied to the change in score of each group. As a result, a significant change was seen in the score of the regular Paro group ($p < 0.05$).

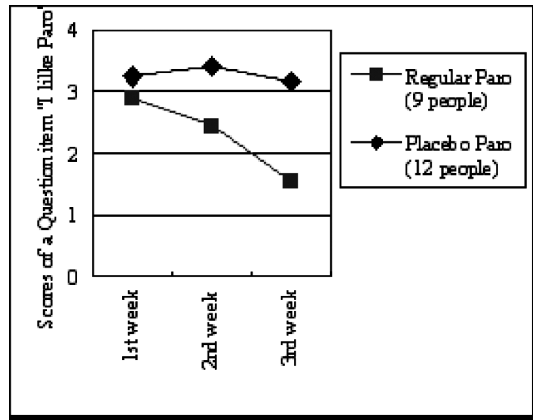


Figure 10. Average scores of a question item “I like Paro” to elderly people

5. Discussion

The effects of Paro on elderly people staying in a health service facility for the aged were investigated. The effects of the regular Paro were then compared with those of a placebo Paro. Against expectations, the face scale scores for regular and placebo Paro groups improved, and their standard scores of Depression-Dejection of POMS decreased after introduction of Paro. From these results, regular and placebo Paro both improved elderly people’s moods. In particular, Paro was effective in combating their depression.

Additionally, familiarity with Paro was investigated through questionnaires. The results were interesting. For the question item “I like Paro”, the average score of the regular Paro group decreased. On the other hand, the average score of the placebo Paro group maintained a high value. Thus, the subjects of the placebo Paro group did not lose interest in the placebo Paro, while the regular Paro group’s interest in their Paro decreased.

Before the experiment, it had been expected that people would lose interest in the placebo Paro, because its reaction was very simple. However, this expectation was wrong. Subjects in the placebo Paro group kept interacting with their Paro, and they did not notice that placebo Paro’s reaction was simple.

From these results, following 3 questions arise:

- Q1. Was the placebo Paro really liked more than the regular Paro by subjects?
- Q2. Why didn’t subjects lose interest in the placebo Paro?
- Q3. Why did the regular Paro group’s interest in their Paro decrease?

Considering Q1, there were several differences between the subjects in the regular and placebo Paro groups, rendering it difficult to make a simple comparison between the regular and placebo Paro. For example, there was one man who made people excited, and the placebo Paro group was more independent than the regular Paro group. However, the subjects could not be randomized because of limitations in the institution.

In the case of Q2, the following 2 reasons are considered:

1. It was difficult for subjects to notice that placebo Paro's reaction was repetitive. An aged person's cognitive capabilities are inferior to those of a normal functioning adult. In addition, subjects interacted with Paro two or more people at the same time. Therefore, each subject's interaction time with Paro was not long enough to notice that its reaction was repetitive.
2. The reaction to crying and raising its head had special meanings. Some subjects said "good boy" when Paro raised its head. They felt that Paro answered their call.

In the case of Q3, it is considered that people might have felt that the regular Paro was impolite, because its reactions to stimuli were too varied (including ignoring them).

In order to clarify these points, experiments will be carried out using a larger number of Paros, to compare the effects of Paro, those of the placebo Paro and those of alternative placebo Paros (e.g. swinging its head to right and left in response to stimuli) on the same subject.

In this research, questionnaires and POMS were used because they can accurately measure six mood states. However, the POMS contained many items, and some subjects refused to answer them over time. Simpler questionnaires will be devised to measure the moods of elderly people.

6. Conclusion

Seal-type "mental commitment robots" Paro were applied to robot-assisted activity for elderly people at a health service facility for the aged. The experiment was carried out over 4 weeks in total. The effects of the regular Paro were then compared with those of a placebo Paro. The results show that interaction with both regular and placebo Paros has both psychological and social effects on elderly people. Physiologically, urinary tests were used to establish that robot-assisted activity decreased the stress reaction in the elderly clients. The details are described in (Saito *et al.* 2002b).

Further experiments and research will be carried out under different conditions and situations. Moreover, an investigation will be conducted into the relationship between the functions of a mental commitment robot and its effects on elderly people participating in robot-assisted activity.

References

- Baun, M. M., Bergstrom, N., Langston, N. F. & Thoma, L. (1984). Physiological Effects of Human/Companion Animal Bonding. *Nursing Research*, Vol. 33, No. 3, 126–129.
- Delta Society (1991). Animal-Assisted Therapy and Crack Babies: a New Frontier, *Delta Society Newsletter*, Vol. 1, No. 2.
- McNair, D. M., Lorr, M. & Droppleman, L. F. (1992). *Profile of Mood States*. San Diego: Educational and Industrial Testing Service.
- Friedmann, E., Katcher, A. H. & Lynch, J. J. (1980). Animal Companions and One-year Survival of Patients after Discharge from a Coronary Care Unit. *Public Health Reports*, Vol. 95, No. 4, 307–312.
- Fujita, M. & Kitano, H. (1998). An Development of an Autonomous Quadruped Robot for Robot Entertainment. *Autonomous Robots*, Vol. 5, 7–18.
- Gammonley, J. & Yates, J. (1991). Pet Projects Animal Assisted Therapy in Nursing Homes. *Journal of Gerontological Nursing*, Vol. 17, No. 1, 12–15.
- Garrity, T., Stallones, F. L. & Marx, M. B. (1989). Pet Ownership and the Elderly. *Anthrozoos*, Vol. 3, No. 1, 35–44.
- Hart, L. A., Hart, B. L. & Bergin, B. (1987). Socializing Effects of Service Dogs for People with Disabilities. *Anthrozoos*, Vol. 1, No. 1, 41–44.
- Haladay, J. (1989). Animal Assisted Therapy for PWAs – Bringing a Sense of Connection. *AIDS Patient Care*, 38–39.
- Kale, M. (1992). Kids & Animals. *Inter Actions*, Vol. 10, No. 3, 17–21.
- Lago, D., Delaney, M., Miller, M. & Grill, C. (1989). Companion Animals, Attitudes Toward Pets, and Health Outcomes Among the Elderly: A Long-Term Follow-up. *Anthrozoos*, Vol. 3, No. 1, 25–34.
- Lorish, C. D. & Maisiak, R. (1986). The Face Scale: A Brief, Nonverbal Method for Assessing Patient Mood. *Arthritis and Rheumatism*, Vol. 29, No. 7, 906–909.
- McNair, D. M., Lorr, M. & Droppleman, L. F. (1992). *Profile of Mood States manual* (rev. ed.). San Diego: Educational and Industrial Testing Service.
- Maslach, C. (1976). Burned-out. *Human Behavior*, Vol. 5, No. 9, 16–22.
- Petroski, H. (1996). *Invention by Design*. Harvard University Press.
- Saito, T., Shibata, T., Wada, K. & Tanie, K. (2002a). Examination of Change of Stress Reaction by Urinary Tests of Elderly before and after Introduction of Mental Commit Robot to an Elderly Institution. *Proc. of the 7th Int. Symp. on AROB*, Vol. 1, 316–319.
- Saito, T., Shibata, T., Wada, K. & Tanie, K. (2002b). Change of Stress Reaction by Introduction of Mental Commit Robot to a Health Services Facility for the Aged. *Proc. of Joint 1st Int. Conf. on SCIS and ISIS*, paper number 23Q1-5.
- Sánchez, E., Shibata, T. & Zadeh, L. (1997). *Perspectives of Fuzzy Logic and Genetic Algorithms*. Scientific World Co. Ltd.

- Shibata, T. (2004). An Overview of Human Interactive Robots for Psychological Enrichment. *Proc. of the IEEE*, 92(11), 1749–1758.
- Shibata, T., Abe, T., Tanie, K. & Nose, M. (1996a). Skill Based Motion Planning in Hierarchical Intelligent Control of a Redundant Manipulator. *Robotics and Autonomous Systems*, 18, 65–73.
- Shibata, T., Inoue, K. & Irie, R. (1996b). Emotional Robot for Intelligent System – Artificial Emotional Creature Project. *Proc. of 5th IEEE Int'l Workshop on Robot and Human Interactive Communication*, ROMAN 1996 (pp. 466–471). IEEE Press.
- Shibata, T. & Irie, R. (1997). Artificial Emotional Creature for Human-Robot Interaction – A New Direction for Intelligent System. *Proc. of the IEEE/ASME Int'l Conf. on AIM'97 (Jun. 1997)* paper number 47 and 6 pages in CD-ROM Proc.
- Shibata, T., Tashima, T. & Tanie, K. (1999). Emergence of Emotional Behavior through Physical Interaction between Human and Robot. *Proc. of the 1999 IEEE Int'l Conf. on Robotics and Automation*, ICRA 1999 (pp. 2868–2873). IEEE Press.
- Shibata, T. & Tanie, K. (2000). Influence of A-Priori Knowledge in Subjective Interpretation and Evaluation by Short-Term Interaction with Mental Commit Robot. *Proc. of the IEEE Int'l Conf. On Intelligent Robot and Systems*, IROS 2000 (pp. 169–172). IEEE Press.
- Shibata, T., Mitsui, T., Wada, K., Touda, A., Kumasaka, T., Tagami, K. & Tanie, K. (2001a). Mental Commit Robot and its Application to Therapy of Children. In B. Siciliano (Ed.), *Proc. of the IEEE/ASME Intl. Conf. on Advanced Intelligent Mechatronics*, AIM'01 (pp. 1053–1058). July 8–12, 2001, Como, Italy.
- Shibata, T., Wada, K., Saito, T. & Tanie, K. (2001b). Robot Assisted Activity for Senior People at Day Service Center. In *Proc. of the First Int'l. Conference on Information Technology in Mechatronics*, ITM'01 (pp. 71–76). October 1–6, 2001, Istanbul and Capadocia, Turkey.
- United Nations (1998). *World Population Prospects: The 1998 Revision*.
- Wada, K., Shibata, T., Saito, T. & Tanie, K. (2004). Effects of Robot Assisted Activity for Elderly People and Nurses at a Day Service Center. *Proc. of the IEEE*, 92, 11, 1780–1788.
- Wada, K., Shibata, T., Saito, T., Sakamoto, K. & Tanie, K. (2005a). Psychological and Social Effects of One Year Robot Assisted Activity on Elderly People at a Health Service Facility for the Aged. In *Proc. of the 2005 IEEE Intl. Conference on Robotics and Automation*, ICRA 2005 (pp. 2796–2801). IEEE Press.
- Wada, K., Shibata, T., Musha, T. & Kimura, S. (2005b). “Effects of Robot Therapy for Demented Patients Evaluated by EEG.” *Proc. of the IEEE/RSJ Int'l Conf. on IROS*, 2205–2210.
- Yokoyama, A. (2002). “The Possibility of the Psychiatric Treatment with a Robot as an Intervention – From the Viewpoint of Animal Therapy.” *Proc. of Joint 1st Int. Conf. on SCIS and ISIS*, paper number 23Q1-1.

Designing avatars for social interactions

Marc Fabri, David J. Moore and Dave J. Hobbs

1. Introduction

Natural human communication is based on speech, facial expressions, body posture and gestures. While speech is an obvious instrument for mediating our thoughts and ideas, social intercourse also depends heavily on the expressions and movements of the body (Morris *et al.* 1979). Such socio-emotional content is vital for building relationships that go beyond the purely factual and task-oriented communication usually encountered in a business environment. Indeed, social psychologists argue that more than 65% of the information exchanged during a face-to-face conversation is carried on the nonverbal band (Knapp 1978; Argyle 1988). Further, recent findings in psychology and neurology suggest that emotions are an important factor in decision-making, problem solving, cognition and intelligence in general (Damásio 1994; Picard 1997; Lisetti & Schiano 2002; Kaiser & Wehrle, this volume; Gratch & Marsella 2005). We therefore expect that the inclusion of such communication channels in collaborative virtual environments (CVEs) will be beneficial in some way.

CVEs offer a stimulating and integrated framework for conversation and collaboration. Indeed, it can be argued that CVEs represent a communication technology in their own right due to the highly visual and interactive character of the interface that allows communication and the representation of information in new, innovative ways. Users are likely to be actively engaged in interaction with the virtual world and with other inhabitants.

In CVEs, inhabitants are usually represented by humanoid embodiments, referred to as “avatars”. Since the avatar is part of the perceived environment and at the same time represents the user that is perceiving (Slater & Wilbur 1997), inhabitants potentially develop a strong sense of mutual awareness. The avatar can provide direct feedback to the other inhabitants about a user’s actions, degree of attention and interactive abilities, and therefore it becomes an effective interaction device.

Although research in the field of CVEs has been proceeding for some time now, the representation of avatars in many systems is still relatively simple and

rudimentary. Attempts have been made to improve avatar appearance and behavioural qualities (cf. Manninen & Kujanpää 2002; Garau 2003; Vinayagamoorthy *et al.* 2004). An area that has received less attention though is the emotional expressiveness of avatars. Interaction in CVEs has been notoriously poor in terms of emotional cues.

As far back as the late 1990s, Dumas and colleagues (1998) pointed out the need for sophisticated ways to reflect emotions in virtual embodiments. Thalmann (2001) saw a direct relationship between the expressive abilities of a user's representation and their ability to interact with the environment and with each other. Slater and colleagues (2000) observed that even avatars with rather primitive expressive abilities could cause strong emotional responses in people using a CVE system. It appears that the avatar can readily take on a personal role, potentially becoming a genuine representation of the underlying individual – not only visually but also in a social context.

Potential applications for CVE systems are all areas where people cannot come together physically but wish to discuss, collaborate on, or even dispute certain matters. We are in particular concerned with the use of CVE technology in Distance Learning systems (see Fabri & Gerhard 2000, for a detailed discussion). Interaction between those involved in the learning process is important for mutual reflection on actions and problem solutions, motivation and stimulation (Laurillard 1993). More specifically, the ability to show emotions, empathy and understanding through the face and body is central to ensuring the quality of tutor-learner interaction (Knapp 1978; Cooper *et al.* 2000). This is reflected in the decision by the UK Department for Education and Skills' to introduce "emotional literacy" as a topic in primary and secondary schools (DfES 2005). It is hoped that pupils learn to judge their skills, and those of their classmates, in five areas: self-awareness, empathy, managing feelings, self-motivation and social interaction.

We argue, then, that CVE technology is a potentially very powerful means of facilitating communication between people working at a distance. However, little is known about how – or indeed whether – emotions can effectively be transmitted through the medium of CVE. This chapter outlines our developmental and experimental work designed to address this gap in our knowledge.

2. The expression of emotion

Various channels are available to express emotion: voice, the face, gaze, gesture, or posture. Of the non-verbal channels, the face is the most immediate indicator for the emotional state of a person (Ekman & Friesen 1975a). While our work focuses on the face and facial expressions of emotion, this book gives a wider outlook on

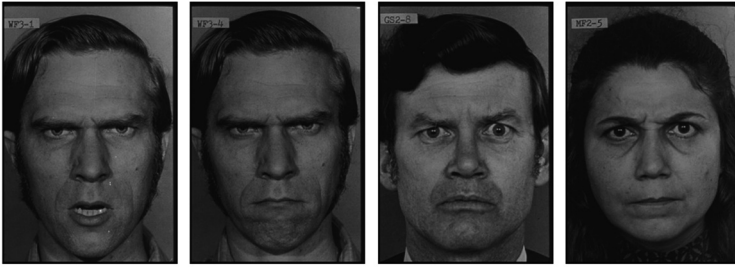


Figure 1. Variations of anger; photographs from (Ekman & Friesen 1975b), courtesy of Paul Ekman

how non-verbal channels can be utilised to create expressive characters – in this volume, see for example Coulson’s chapter on expressing emotion through body movement, and the work on eye gaze in Gillies, Ballin and Dodgson’s chapter.

As socially active humans, we usually have an informal understanding of what emotion is and what different emotions there are. There is also a formal research tradition that has investigated the nature of emotion systematically. Major figures in scientific research have contributed to this investigation, most notably the philosopher Rene Descartes (1641), biologist Charles Darwin (1872), and more recently the psychologist Paul Ekman.

Ekman and colleagues (1972) found that there are six universal facial expressions corresponding to the following emotions: surprise, anger, fear, happiness, disgust/contempt, and sadness. The categorisation is widely accepted, and considerable research has shown that these basic emotions have a clear meaning across cultures (Zebrowitz 1997; Ekman 1999). Indeed, it is held that expression and, to some extent, recognition of these six emotions has an innate basis. Figure 1 shows variations of the anger category.

This naturally developed skill to “read” and indeed generate facial expressions is, we argue, highly beneficial to communication in CVEs. We believe that the expressive virtual face of an interlocutor’s avatar can aid the communication process and provide information that would otherwise be difficult to mediate.

To comprehensively describe the visible muscle movement in the face, Ekman and Friesen (1975a, 1978) developed the Facial Action Coding System (FACS) based on highly detailed anatomical studies of human faces. A facial expression is a high-level description of facial motions and can be decomposed into certain muscular activities (e.g. relaxation or contraction) called Action Units (AUs). FACS identifies 58 AUs, which separately or in various combinations are capable of characterising any human expression. An AU corresponds to an action produced by one or a group of related muscles. Action Unit 7, for example, is the *lid-tightener*, tightening the eyelids and thereby narrowing the eye opening.

3. Modelling expressive faces

Platt and Badler (1981) developed the first muscle-based model of an animated face using geometric deformation operators to control a large number of muscle units. Parke (1982), and later Terzopoulos and Waters (1993) further developed this by modelling the anatomical nature of facial muscles and the elastic nature of human skin. Improved versions of muscle- and skin-based models are in use today.

However, this comes at a price: muscle-based animation models are generally complex and computationally intensive. For that reason they are not widely used in real-time animation (Bui *et al.* 2003). The approach we have chosen, therefore, is feature-based and less complex than a realistic simulation of real-life physiology. This, we argue, is sufficient and in fact preferable, as it allows us at the same time to establish what the most distinctive and essential features of a facial expression are. Furthermore, we argue that it is not necessary, and may indeed be counter-productive, to assume that a “good” avatar has to be a realistic and accurate representation of real world physiognomy. On the basis of experimental studies on avatar responsiveness, visual quality and gaze, Garau (2003) found that full realism is not essential for effective social interaction. Early evidence suggested that approaches aiming to reproduce the human physics in detail may in fact be wasteful (Benford *et al.* 1995). Bailenson and colleagues (2005) consider the development of behavioural realism (as opposed to visual realism) as the far worthier goal on the grounds that this taps into people’s innate and culturally fostered abilities to understand meaning from non-verbal signals, e.g. facial expressions.

Researchers investigating the effect of varying degrees of realism in virtual characters often refer to the *Uncanny Valley* effect. The term was first coined by Mori (1974) and originally intended to measure human psychological reaction to the anthropomorphism of robots (see Figure 2, adapted from Reichardt 1978). When plotting human reaction against robot anthropomorphism, the curve initially shows a steady upward trend, continuing until the robot reaches reasonably human quality. However, it then plunges down dramatically: A *nearly* human robot is considered irritating and repulsive. The curve only rises again once the robot reaches complete resemblance with humans. Likewise, we postulate that human reaction to avatars is similarly characterised by an *uncanny valley*. An avatar designed to suspend disbelief that is only *nearly* realistic may be equally confusing and not be accepted or it might even be considered repulsive.

In any event, Hindmarsh and colleagues (2001) suggest that even with full realism and perceptual capabilities of physical human bodies in virtual space, opportunities for employing inventive and evocative ways of expression would be lost if the focus were merely on simulating the real world with its rules, habits

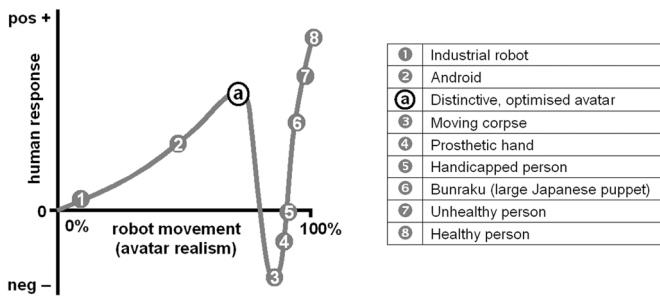


Figure 2. The Uncanny Valley (adapted from Reichardt 1978)

and limitations. Donath (2002) warns that because the face is so highly expressive and humans are so adept in reading (into) it, any level of detail in 3D facial rendering could potentially provoke the interpretation of various social messages. If these are unintentional, the face will arguably be hindering rather than helping communication. It may be more appropriate, and indeed more supportive to perception and cognition, to represent issues in simple or unusual ways. There is evidence that particularly distinctive faces can convey emotion more efficiently than normal faces (Zebrowitz 1997; Bartneck 2002), a detail regularly employed by caricaturists.

To summarise, rather than simulating the real world accurately, we aim to take advantage of the innate cognitive abilities to perceive, recognise and interpret physiognomic clues that humans seem to have. With regard to avatar expressiveness and the uncanny valley, we are targeting the first summit of the curve (Figure 2) where human emotional response is maximised while employing a relatively simple avatar model.

Given all this, in order to realise such an approach in our avatar work, we developed an animated virtual head with a limited number of controllable features. It is based on H-Anim (2008), the specification proposed by an international panel that develops the Virtual Reality Modeling Language (VRML). H-Anim specifies seven control parameters: *left/right eyeball*, *left/right eyebrow*, *left/right upper eyelid*, and *temporomandibular* (for moving the jaw). During our investigation it became evident that this specification was insufficient for the variety of expressions required, in particular in the lower face area. Consequently, control parameters were added and additional features were derived from, and closely mapped to, FACS Action Units. Figure 3 shows the controllable parameters of the virtual head, modelled in VRML.

It is clear that these parameters alone do not allow representation of all facial expressions that are possible in real life. However, we do not need to reproduce the entire set of Action Units to achieve the level of detail envisaged for the

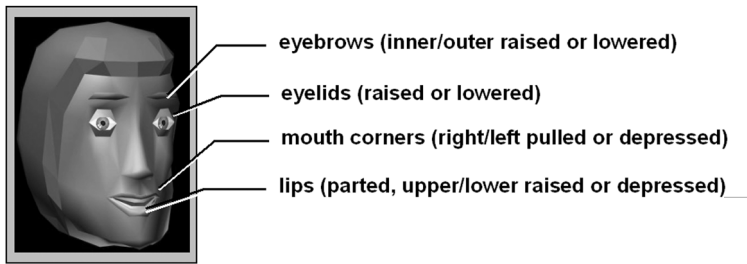


Figure 3. Controllable features of the virtual head modeled in VRML

current face model. The human perceptual system can recognise physiognomic clues, in particular facial expressions, from very few visual stimuli (Dittrich 1993) and our head model is designed to display precisely these distinctive facial clues. In fact, reducing the number of relevant Action Units is not uncommon practice for simple facial animation models (Yacoob & Davis 1994; Parke & Waters 1996; Spencer-Smith *et al.* 2002) and this study uses a subset of eleven Action Units, listed in Table 1.

We regard the virtual head with its limited but human-like expressive abilities as a potentially effective and efficient means to convey emotions in virtual environments. Furthermore, we consider the reduced set of AUs and the resulting facial animation control parameters as being potentially sufficient to express, in a readily recognisable manner, the six universal facial expressions of emotion and the neutral expression. For example, Figure 4 shows the same variations of anger as Figure 1, but now depicted by the virtual head.

Table 1. Reduced set of Action Units employed in our work

AU	Facial Action Code	Muscular Basis
1	Inner Brow Raiser	Frontalis, Pars Medialis
2	Outer Brow Raiser	Frontalis, Pars Lateralis
4	Brow Lowerer	Depressor Glabellae, Depressor Supercilli, Corrugator
5	Upper Lid Raiser	Levator Palpebrae Superioris
7	Lid Tightener	Orbicularis Oculi, Pars Palpebralis
10	Upper Lip Raiser	Levator Labii Superioris, Caput Infraorbitalis
12	Lip Corner Puller	Zygomatic Major
15	Lip Corner Depressor	Triangularis
17	Chin Raiser	Mentalis
25	Lips Part	Depressor Labii, Relaxation of Ment. or Orbicularis Oris
26	Jaw Drop (mouth only)	Maseter, Relaxation of Temporal & Internal Preygooids

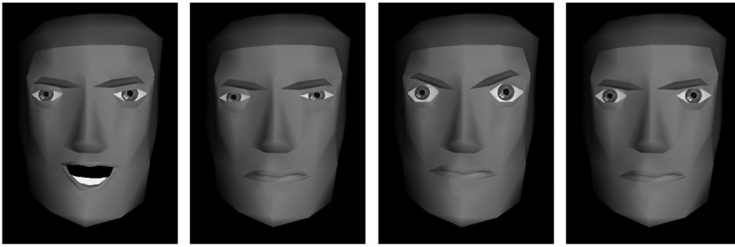


Figure 4. Variations of anger (virtual head model)

4. Experimental study

An experiment was carried out to test the claims argued for in the previous section. Screenshots of the virtual head depicting variations of emotions were shown to participants. Participants were also to look at photographs of people showing corresponding expressions, taken from the “Pictures of Facial Affect” databank (Ekman & Friesen 1975b). The task was to assign each expression to one of the aforementioned categories – i.e. *surprise*, *anger*, *fear*, *happiness*, *disgust/contempt* and *sadness*. Two additional categories, “*Other...*” and “*Don’t know*”, were also available for selection by participants. An interactive application was built to record and compile participants’ responses automatically during the study. The user interface for the recognition task is shown in Figure 5.

A total of 29 participants (17 female and 12 male) with an average age of 30 (ranging from 22 to 51 years old) took part in the experiment. None of them had classified facial expressions or used FACS before. None of the participants worked in facial animation, although some were familiar with 3D modelling techniques in general. In an attempt to equalise possible individual differences in recognition skills we chose a repeated measures design, i.e. participants constituted their own control group: Any given participant would be asked to consider and categorise both photographic and avatar representations of emotions. Each participant was shown 28 photographs and 28 corresponding virtual head images mixed together in a randomly generated order that was the same for all participants. Each of the six emotion categories was represented in 4 variations, together with 4 variations of the neutral face. The variations were defined not by intensity, but by differences in expression of the same emotion. The controllable parameters of the virtual head were adjusted to correspond with the natural faces. All virtual head images depicted the same male model throughout, whereas the photographs showed several people, expressing a varying number of emotions (20 persons were male, 8 female). Photographs were selected from the face databank solely based on their distinctness, i.e. only those that attracted high recognition rates in the original



Figure 5. Recognition screen used in our experimental study

studies carried out by Ekman and Friesen were considered. This was believed to be the most appropriate method, aiming to avoid the introduction of factors that would potentially disturb results, such as gender, age or ethnicity.

5. Results

Statistical analysis showed that recognition accuracy (or distinctness as labelled by Bartneck 2002) for photographs (78.6% overall) is significantly higher than those for virtual heads (62.2%). A Mann-Whitney test at significance level 0.01 confirmed this. Recognition rates vary across emotion categories as well as between the two types of stimuli, as illustrated clearly in Figure 6.

Interestingly, participants who achieved better results did so homogenously between virtual heads and photographs. Lower scoring participants were more likely to fail recognising virtual heads rather than photographs. A closer look at recognition rates of particular emotions reveals that all but *disgust* have at least one photograph/virtual head pair with comparably high results, i.e. recognition was as successful with the virtual head as it was with the directly corresponding photograph. Figure 7 shows recognition results for the highest scoring virtual head per category (depicted in Figure 8).

We conducted a detailed analysis of the errors participants made when asked to categorise facial expressions, shown in Table 2. Rows give per cent occurrence of each response. Confusion values above 10% are shaded light grey, above 20% dark grey, above 30% black.

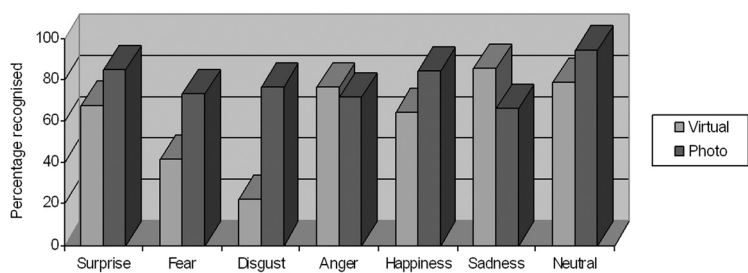


Figure 6. Summary of recognition rates

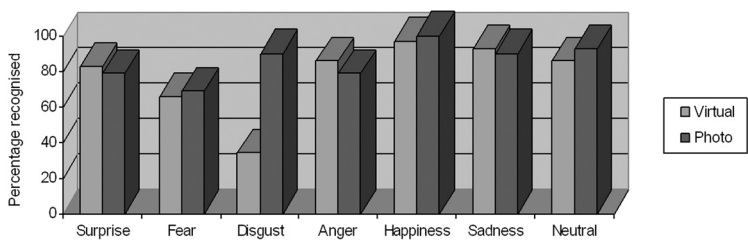


Figure 7. Recognition rates for selected images

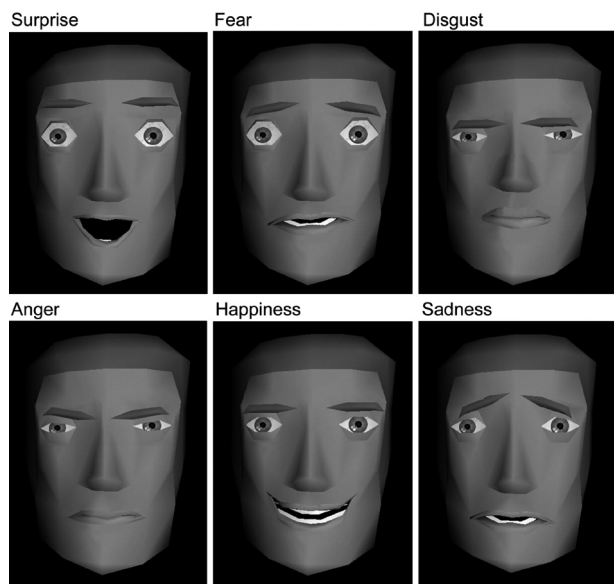


Figure 8. Most distinctive expression in each category

Table 2. Categorisation error matrix

Emotion category	Response [virtual / photographs]							Other/ Don't know
	Surprise	Fear	Disgust	Anger	Happiness	Sadness	Neutral	
Surprise	.67 / .85	.06 / .07	.00 / .00	.00 / .01	.23 / .00	.00 / .00	.01 / .00	.03 / .08
Fear	.15 / .19	.41 / .73	.00 / .04	.30 / .00	.03 / .00	.03 / .00	.02 / .00	.06 / .03
Disgust	.01 / .02	.02 / .00	.22 / .77	.39 / .14	.01 / .00	.04 / .00	.10 / .01	.21 / .07
Anger	.03 / .04	.00 / .04	.00 / .03	.77 / .72	.02 / .00	.03 / .03	.11 / .05	.05 / .09
Happiness	.01 / .00	.01 / .00	.01 / .00	.01 / .00	.64 / .84	.03 / .00	.26 / .15	.04 / .02
Sadness	.06 / .00	.09 / .10	.00 / .00	.00 / .01	.01 / .01	.85 / .66	.03 / .09	.01 / .07
Neutral	.03 / .00	.03 / .00	.01 / .00	.00 / .01	.00 / .02	.11 / .01	.78 / .94	.04 / .02

6. Discussion

6.1 Disgust

The majority of errors were made in the category disgust, an emotion frequently confused with anger. When examining results for virtual heads only, anger was picked nearly twice as often (39%) as disgust (22%).

Disgust is typically shown in the mouth and nose area (Ekman & Friesen 1975a). Although our model features a slightly raised lip (AU10), there is no movement of the nose. Spencer-Smith *et al.* (2002), using a similar setup to investigate intensity of emotions, also found disgust often confused with anger and difficult to depict without well-defined morph targets in the nasal area. Bartneck (2002) observed a very similar effect when studying recognition of facial expressions simplified to merely a few lines and dots. He concluded that the lack of clues around the nose in his model was the likely cause. This strongly suggests that to improve distinctiveness of the disgust expression in a real-time animated model, the nose should be included in the animation, as should the relevant action unit AU9 which is responsible for “nose wrinkling”. Given this, we have now developed an animated model of the virtual head that is capable of lifting and wrinkling the nose to express disgust.

6.2 Fear and surprise

The error matrix further reveals that fear was often mistaken for surprise, a tendency documented in the psychology literature (Ekman 1999) as well as in studies on expressive robots (Cañamero & Fredslund 2001; Bartneck 2002) and animated characters (Spencer-Smith *et al.* 2002). Incidentally, the experience and therefore expression of fear and surprise often happens simultaneously, e.g. when fear is felt

suddenly due to an unexpected threat (Ekman & Friesen 1975a). The appearance of fear and surprise is also similar, with fear generally producing a more tense facial expression. However, Ekman and Friesen (1975a) see fear differing from surprise in three ways:

1. Whilst surprise is not necessarily pleasant or unpleasant, even mild fear is unpleasant.
2. Something familiar can induce fear, but hardly surprise (for example a visit to the dentist).
3. Whilst surprise usually disappears as soon as it is clear what the surprising event was, fear can last much longer, even when the nature of the event is fully known.

These indicators enable people to distinguish between fear and surprise in others. All three have to do with context and timing of the fear-inspiring event – factors that are not perceivable from still images. In accordance with this, Poggi and Pelachaud (2000) found that emotional information is not only contained in the facial expression itself, but also in the performatives of a communicative act: suggesting, warning, ordering, imploring, approving and praising.

Bartneck (2002) found that recognition rates for still images of facial expressions were higher when shown in a game context, compared to images shown out of context, as was the case in our current study. In other words, the inherent meaning and subjective interpretation of an emotional expression depend on the situation in which it is shown. This strongly suggests that when using emotionally expressive avatars in CVEs where there will be both the context and triggers for the emotion display, recognition rates will be higher than those observed in the current study.

6.3 Fear and anger

The relationship between fear and anger is similar to that between fear and surprise. Both can occur simultaneously, and their appearance often blends. What is striking is that all confusions between the two emotions in our data concerned virtual faces. This may suggest that the fear category contained some particularly unsuitable examples of modelled facial expressions. An examination of results showed that one virtual head expression of fear in particular, shown in Figure 9, was regularly mistaken for anger.

The expression on the left of Figure 9 displays characteristic fear eyes. The lower eyelid is visibly drawn up and appears to be very tensed. Both eyebrows are slightly raised and drawn together. The lower area of the face also shows charac-

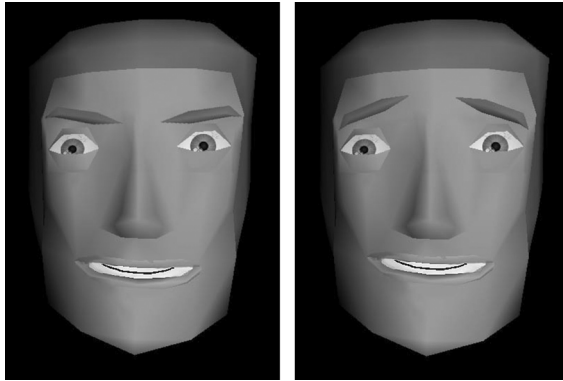


Figure 9. Unsuitable fear expression (left) and alternative fear expression (right)

teristics of fear, namely a slightly opened mouth with stretched lips that are drawn together. In contrast, an angry mouth has the lips either pressed firmly together or open in a squarish shape, as if to shout (Ekman & Friesen 1975a). However, despite these seemingly characteristic clues, 18 out of 29 times this expression was categorised as anger. Presumably, the main reason for this is the eyebrow: In anger, as in fear, eyebrows can be drawn together. But unlike the fearful face which shows raised eyebrows, angry faces typically feature a lowered brow. Consider the expression on the right in Figure 9, which is identical to that on the left apart from the eyebrows. They are now raised and arched, a detail that changes the facial expression significantly, making it less ambiguous and distinctively fearful. Incidentally, no participant categorised expression 9 (right) as anger, and it had the second-highest recognition rate of all fear expressions.

6.4 Choice of emotion categories

Results also indicate that limiting the number of categories might have had a negative effect on the recognition scores because of the number of “Other” and “Don’t know” responses. This could probably be avoided in the future by allowing more categories, or alternatively by offering a range of suitable descriptions for an emotion category (e.g. joy, cheerfulness and delight, to complement happiness). On the other hand, Cañamero and Fredslund (2001) suggest that the very idea of using words to categorise emotions may cause confusion for emotions that are intuitively very easy to recognise. Further work is necessary to establish how best to capture what individuals perceive and recognise, and how this is documented.

Overall, it has to be noted that many of the terms suggested by participants as their “Other...” choice were actually very similar to the emotion category expected, confirming that the facial expressions in those cases were not necessarily badly depicted. This does however highlight the importance of having a well-defined vocabulary when investigating emotions – a problem that is not new to the research community and that has been discussed at length over the years (see Ekman *et al.* 1972 for an early comparison of dimensions vs. categories; Zebrowitz 1997; Lisetti & Schiano 2000; also Kaiser & Wehrle’s chapter in this volume).

7. Conclusions and further work

The experimental work discussed in this chapter suggests that, when applying the FACS model to virtual face representations, emotions can effectively be visualised with a very limited number of facial features. For example, in respect of the “top scoring” virtual heads, emotion recognition rates are, with the exception of the disgust emotion, comparable to those of their corresponding real-life photographs. These top-scoring expressions are exemplar models for which detailed AU scoring is available.

Although it remains to be corroborated through further studies, it is believed that such simple, pure emotional expressions could fulfil a useful role in displaying explicit, intended communicative acts which can therefore help interaction in a CVE. They can provide a basis for emotionally enriched CVEs, and hence for the benefits of such technology being used, for example, within distance learning as argued for earlier.

It is noted that such pure forms of emotion are not generally seen in real life, as many expressions occurring in face-to-face communication between humans are unintended or automatic reactions. They are often caused by a complex interaction of several simultaneous emotions. This is vividly illustrated in Picard’s (1997) example of a Marathon runner after the finish, discussed by Carofiglio, de Rosi and Grassano in their chapter on Mixed Emotion Activation of this volume. With regards to our own work, a different expression control mechanism such as capturing an individual’s facial expression directly by using a camera, for example, may allow varying intensities and blends of expressions to be recognised and modelled onto avatar faces. However, this study has deliberately opted for an avatar that can express clearly, and hopefully unambiguously, what a controlling individual exactly wants it to express.

The data analysis also suggests that the approach advocated earlier is not guaranteed to work for all expressions, or all variations of a particular emotion category. Further evidence is supplied in the post-experiment questionnaire data.

Two participants, for example, noted that on several occasions the virtual face expression was not distinctive enough, and others observed that the virtual head showed no lines or wrinkles and that recognition might have been easier with these visual cues. Not surprisingly, the issues arising here are similar to those identified by real-life social psychology.

Firstly, no categorisation system can ever be complete. Although accepted categories exist, emotions can vary in intensity and inevitably there is a subjective element to recognition. When modelling or animating facial features, such ambiguity in interpretation can be minimised by focussing on, and emphasising the most distinctive visual cues of a particular emotion.

Secondly, context plays a crucial role in emotion expression and recognition. Effective, accurate mediation of emotion is closely linked with the situation and other, related communicative signals. A reliable interpretation of facial expressions without taking cognisance of the context in which they are displayed is often not possible. One would expect, therefore, that recognition of avatar representations of emotion will be higher when contextualised. This assumption requires empirical investigation, however, and our immediate next step in this research is to carry out controlled experiments to address this. A further contextual issue concerns culture. Although emotions exist universally, there can be cultural differences concerning when and how emotions are displayed (Zebrowitz 1997). It appears that people in various cultures differ in what they have been taught about managing or controlling their facial expression of emotion. Ekman and Friesen (1975a) call these cultural norms “display rules”. How such cultural differences might play themselves out in a virtual world is an important open question, which we plan shortly to address.

Thirdly, further work on emotion recognition in a real-time virtual reality setting has to consider the effects timing and intensity have on emotion display and interpretation. Already there is evidence that recognition accuracy of dynamic, animated faces is considerably higher than for static faces, in particular when the emotion displayed is of low intensity (Spencer-Smith *et al.* 2002).

Finally, the authors propose to explore a potential application of emotionally expressive avatars in a specialist field, namely as part of the developmental training given to people diagnosed with autism. One aspect of autism is a so-called “theory of mind deficit” (Howlin *et al.* 1999) – people with autism may have difficulty in understanding mental states and emotions and in ascribing them to themselves or to others. We argue that a CVE incorporating avatars capable of displaying emotions such as those used in this experiment might be of value in developing in autistic people the ability to learn how to recognise, interpret, and act appropriately on emotions displayed by others. Further, autistic people may be able to use such a CVE in a prosthetic role. Parsons and Mitchell (2002) argue that

because interactions via CVE tend to be slower than face-to-face interactions, it provides users with autism time to think of alternative ways of dealing with a particular situation. The technology may therefore provide a means by which people with autism can communicate with others, and this may help them circumvent to some extent their social and communication impairment and sense of isolation (Moore *et al.* 2005). Such CVE tools may also be less threatening to people with autism than face-to-face communication, and thereby help avoiding many of the potential pitfalls of the real world (Parsons *et al.* 2005). There is a danger, however, that users may find the non-social nature of computer-based tasks so appealing that they become overly reliant on the technology (Howlin *et al.* 1999), a concern also raised by Parsons and colleagues (2004).

Thus, there is scope for much additional research concerning the value and importance within social interactions of avatars capable of expressing a range of emotions in ways that are believable and acceptable to humans. The authors contend that the experimental evidence discussed in this chapter can inform and provide a good foundation for this work.

Acknowledgements

We would like to thank Dr. Paul Ekman for giving permission to reproduce selected PFA Photographs. Original virtual head geometry copyright Geometrek.

References

- Argyle, M. (1988). *Bodily Communication*. New York: Methuen.
- Bailenson, J., Yee, N., Merget, D. & Schroeder, R. (2005). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, non-verbal disclosure, emotion recognition and copresence in dyadic interaction. In M. Slater (Ed.), *Proceedings of the 8th Intl. Workshop on Presence, PRESENCE 2005*, Department of Computer Science, UCL (University College London), London, UK, 21–23 September, 2005.
- Barneck, C. (2002). *eMuu – An Embodied Emotional Character for the Ambient Intelligent Home*. PhD thesis. Technische Universiteit Eindhoven. ISBN 9038618875.
- Benford, S., Bowers, J., Fahlén, L., Greenhalgh, C. & Snowdon, D. (1995). User Embodiment in Collaborative Virtual Environments. In I. R. Katz, R. Mack, L. Marks, M. B. Rosson & J. Nielsen (Eds.), *Proc. Conference on Human Factors in Computing Systems (CHI'95)*. New York, NY: ACM Press.
- Cañamero, L. & Fredslund, J. (2001). I show you how I like you – can you read it in my face? *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 31 (5), 454–459.

- Cooper, B., Brna, P. & Martins, A. (2000). Effective Affective in Intelligent Systems: Building on Evidence of Empathy in Teaching and Learning. In A. Paiva (Ed.), *Affective Interactions* (pp. 21–34) [LNAI 1814]. Heidelberg & Berlin: Springer-Verlag.
- Damásio, A. (1994). *Descartes's Error: Emotion, Reason and the Human Brain*. NY: Avon.
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. Oxford University Press (Reprint).
- Descartes, R. (1641). *Meditations and Other Metaphysical Writings*, Penguin (Reprint).
- DfES (2005). Social and emotional aspects of learning, *Primary National Strategy for Behaviour and Learning*, Department for Education and Skills Publication Centre, Nottingham, Ref. DfES0110-2005G.
- Dittrich, W. (1993). Action categories and the perception of biological motion. *Perception*, 22, 15–22.
- Donath, J. (2001). Mediated Faces. In M. Beynon, C. Nehaniv & K. Dautenhahn (Eds.), *Proc. Cognitive Technology: Instruments of Mind. 4th International Conference (CT-2001)*. Coventry, UK, August 6–9, 2001.
- Dumas, C., Saugis, G., Chaillou, C., Degrande, S. & Viaud, M. L. (1998) A 3-D Interface for Cooperative Work. In D. Snowdon & E. Churchill (Eds.), *Proceedings of Collaborative Virtual Environments 1998 (CVE'98)*, University of Manchester, 17–19 June, 1998, Manchester, UK.
- Ekman, P. (1999). Facial Expressions. In T. Dalgleish & M. Power (Eds.) *Handbook of Cognition and Emotion*. New York: John Wiley & Sons.
- Ekman, P. & Friesen, W. (1975a). *Unmasking the Face*. New Jersey: Prentice-Hall.
- Ekman, P. & Friesen, W. (1975b). *Pictures of Facial Affect (Databank on CD-Rom)*. San Francisco University of California, Department of Psychology.
- Ekman, P. & Friesen, W. (1978). *Facial Action Coding System*. Cons. Psychologists Press.
- Ekman, P., Friesen, W. & Ellsworth, P. (1972). *Emotion in the Human Face*. NY: Pergamon.
- Fabri, M. & Gerhard, M. (2000). The Virtual Student. In G. Orange & D. Hobbs (Eds.), *International Perspectives on Tele-Education and Virtual Learning Env.*, Ashgate.
- Garau, M. (2003). *The Impact of Avatar Fidelity on Social Interaction in Virtual Environments*, unpublished PhD Thesis, University College London, UK.
- Gratch, J. & Marsella, S. (2005). Evaluating a computational model of emotion. *Journal of Autonomous Agents and Multiagent Systems*, 11(1), 23–43.
- H-Anim (2008). *Humanoid Animation Specification*, ISO/IEC FCD 19774: 200X, URL <http://www.h-anima.org>
- Hindmarsh, J., Fraser, M., Heath, C. & Benford, S. (2001). Virtually Missing the Point: Configuring CVEs for Object-Focused Interaction. In Churchill, Snowdon and Munro (Eds.), *Collaborative Virtual Environments*. London: Springer Verlag
- Howlin, P., Baron-Cohen, S. & Hadwin, J. (1999). Teaching Children with Autism to Mind-Read, *A Practical Guide for Teachers and Parents*. John Wiley and Sons.
- Knapp, M. (1978). *Nonverbal Communication in Human Interaction*. New York: Holt Rinehart Winston.
- Laurillard, D. (1993). *Rethinking University Teaching*. London: Routledge.
- Lisetti, C. & Schiano, D. (2000). Facial Expression Recognition: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect. *Pragmatics and Cognition*, 8(1), 185–235.

- Manninen, T. & Kujanpää, T. (2002). Non-Verbal Communication Forms in Multi-player Game Sessions. In X. Faulkner, J. Finlay & F. D tienne (Eds.), *People and Computers XVI – Memorable Yet Invisible*. London: BCS Press.
- Moore, D., Cheng, Y., McGrath, P. & Powell, N. J. (2005). Collaborative virtual environment technology for people with autism. *Focus on Autism and Other Developmental Disabilities*, 20(4), 231–243.
- Mori, M. (1974). *The Buddha in the Robot*. Tuttle Publishing.
- Morris, D., Collett, P., Marsh, P. & O’Shaughnessy, M. (1979). *Gestures, their Origin and Distribution*. London: Jonathan Cape.
- Parke, F. (1982). Parameterized modeling for facial animation. *IEEE Computer Graphics and Applications*, 2(9), 61–68.
- Parke, F. & Waters, K. (1996). *Computer Facial Animation*. AK Peters Ltd.
- Parsons, S., Mitchell, P. & Leonard, A. (2005). Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments? *Autism*, 9, 95–117.
- Parsons, S., Mitchell, P. & Leonard, A. (2004). The use and understanding of virtual environments by adolescents with autistic spectrum disorders, in *Journal of Autism and Developmental Disorders*, 34(4), 449–466.
- Parsons, S. & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders, in *Journal of Intellectual Disability Research*, 46, 430–443.
- Picard, R. (1997). *Affective Computing*. Cambridge, MA: MIT Press.
- Platt, S. & Badler, N. (1981). Animating facial expression. *ACM SIGGRAPH Proceedings*. 15(3), 245–252.
- Poggi, I. & Pelachaud, C. (2000). Emotional Meaning and Expression in Animated Faces. In A. Paiva (Ed.), *Affective Interactions* (pp. 182–195) [LNAI 1814]. Heidelberg & Berlin: Springer-Verlag.
- Reichardt, J. (1978). *Robots: Fact, Fiction and Prediction*. London: Thames & Hudson.
- Slater, M., Sadagic, A., Usoh, M. & Schroeder, R. (2000). Small Group Behaviour in a Virtual and Real Environment: A Comparative Study. *Presence*, 9(1), 37–51.
- Slater, M. & Wilbur, S. (1997). Speculations on the Role of Presence in Virtual Environments. *Presence*, 6(6), 603–616.
- Spencer-Smith, J., Innes-Ker, A., Wild, H. & Townsend, J. (2002). Making Faces with Action Unit Morph Targets. In R. Aylett & L. Ca amero (Eds.), *Proceedings of the AISB’02 Symposium on Animating Expressive Characters for Social Interactions*. Imperial College, London, April 4–5, 2002. ISBN 1902956256.
- Thalmann, D. (2001). The Role of Virtual Humans in Virtual Environment Technology and Interfaces. In R. Earnshaw, R. Guedj & J. Vince (Eds.), *Frontiers of Human-Centred Computing, Online Communities and Virtual Environments*. London: Springer.
- Vinayagamoorthy, V., Garau, M., Steed, A. & Slater, M. (2004). An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *The Computer Graphics Forum Journal*, 23(1), 1–11.

Applying socio-psychological concepts of cognitive consistency to negotiation dialog scenarios with embodied conversational characters

Thomas Rist and Markus Schmitt

1. Introduction

A good deal of research in the area of intelligent agent systems is driven by the idea to delegate tasks to an agent (Maes 1994). The accomplishment of delegated tasks often requires a user's agent to get in contact and negotiate with agents of other users. In many cases the user will only be interested in the outcome of such a negotiation. For instance, when sending out a "bargain finder agent" the user may not care about the number of online shops visited or the details of the negotiation process - as long as the agent was able to make a reasonable bargain buy. In this case the result of the task delegation is in the foreground. There are, however, other situations in which the user may have an interest in learning about how a certain negotiation result came about. In human-human negotiations, this is of particular interest in cases where the result of a negotiation cannot be explained on the basis of a solely rational argumentation but only if the social context and the personalities of the negotiating parties are considered as well.

Within the context of the EU project MagiCster we have investigated negotiation scenarios with affective, embodied conversational characters embedded in a social context. Somewhat similar to an arena, users send their delegates (avatars) to a virtual space where the avatars negotiate on behalf of their owners. Both result and process of a negotiation can be displayed to the users in form of a simulation using embodied conversational characters. Dealing with such "Avatar Arenas" requires a framework that covers the modeling of affective agents together with dynamically changing social relationships among them.

Technically speaking, an Avatar Arena is a distributed n:1 client server architecture (see Figure 1). While the server component provides the arena where the negotiation takes place, a client component allows a user to configure and

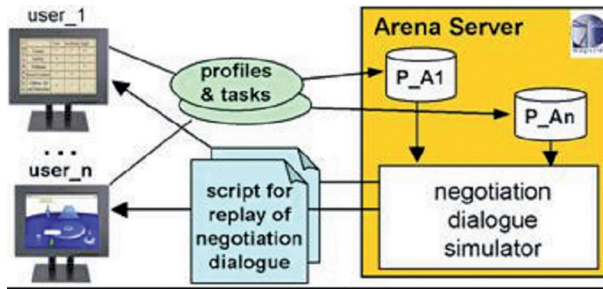


Figure 1. Functional view of Avatar Arena

instruct her/his avatar, and also to observe the negotiation process carried out at the server. To this end, the client receives a generated script of the overall negotiation dialogue for display.

Arena avatars negotiate meeting appointments on behalf of human users. However, we have picked this domain just for the purpose of illustration and do not attempt to make a contribution to meeting planning or appointment scheduling as such. Rather, our research interest is solely in a simulation of the dynamics of social relationships among affective characters during the negotiation dialogues. To this end, Avatar Arena serves as a test-bed to investigate and evaluate mind models of different “cognitive complexity” for the virtual characters that engage in negotiation dialogues. Moreover, our working hypothesis is that an increase of believability in the observable interactions among the characters will indeed require some higher degree of cognitive modeling. Our approach for assessing believability of negotiation dialogues is to show human observers several negotiation dialogues with virtual characters that differ in the number of psychological factors taken into account in a character’s mind model. To this end we first introduce a “basic” Avatar Arena. In this system we only provide our negotiating characters with some domain knowledge about appointment dates and rudimentary conversational skills that enable them to propose meeting dates, and to accept or reject proposals based on their personal calendar entries.

In a second phase we introduce characters that have attitudes concerning already scheduled meeting dates, and new dates to be negotiated. That is, we allow the users to assign importance values to appointments that the avatars should take into account in a negotiation.

Finally, we will consider characters that also have attitudes towards other characters. That is, we allow the users to indicate liking relationships holding between themselves and other users. In the resulting version of Avatar Arena, the avatars consider: (i) their own attitudes towards meeting dates, (ii) attitudes

towards other avatars, and (iii) beliefs about the other avatars' attitudes towards meeting dates.

2. Related work

While there are elaborated approaches for modeling affect in synthetic characters (e.g., see Paiva 2000; Gratch & Marsella 2001; de Rosis *et al.* 2002), relatively little attention has been paid so far to the potential impact of social relationships on a character's behavior. One reason for this might be the fact that it is often less clear what kinds of social relationships should be assumed and modeled when a human user interacts with a synthetic character. When switching to scenarios where multiple characters interact with each others (e.g., André & Rist 2001), however, the need to model social relationships between the involved characters becomes more apparent, and some research groups have already started to account for the social dimension in simulated conversations between animated characters. The work by Prendinger and Ishizuka (2001) deserves mentioning here. In their SCREAM system they explicitly model concepts, such as social distance, social power and threat, in order to enhance the believability of generated dialogues.

Avatar Arena is a multi-character system and thus might be compared to other multi-character systems.

The Agneta and Frida system (Höök *et al.* 1999) incorporates narratives into a Web environment by placing two characters on the user's desktop. These characters watch the user during the browsing process and make comments on the visited Web pages. Dialogue contributions as well as gestures and body movements of Frida and Agneta are pre-authored by a human scriptwriter. Consequently, it depends solely on the skills of the human author whether or not the two characters are recognizable as distinct believable characters. Obviously the scripting approach is not feasible for applications, such as the Avatar Arena, in which characters are configured and instructed independently by different users, and where the characters need to negotiate with each other on behalf of their users.

The Gillbert and George system by Cassell and colleagues automatically generates and animates dialogues between a bank teller and a bank employee with appropriate synchronized speech, intonation, facial expressions, and hand gestures (Cassell *et al.* 1994). However, their focus is on the communicative function of an utterance and not on the personality and the emotions of the single speakers. Furthermore, they do not aim to convey information from different points of view but restrict themselves to a question-answering dialogue between the two animated agents. As entertaining as the generated dialogues may be for a human

observer, they lack some important ingredients that one would find in human-human dialogues dealing with the same task.

Mr. Bengo (Nitta *et al.* 1997) is a system for the resolution of disputes containing three characters: a judge, a prosecutor, and an attorney that is controlled by the user. The prosecutor and the attorney discuss the interpretation of legal rules. Finally, the judge decides who the winner is. The system is noteworthy because it includes a full multimodal interface consisting of components for the recognition and synthesis of speech and facial displays. The virtual characters are able to exhibit some basic emotions, such as anger, sadness, and surprise, by means of facial expressions. However, they neither rely on a solid model for emotion triggering, nor do they exploit other means, such as linguistic style or affective speech, to convey personality or emotions.

Hayes-Roth and colleagues have implemented several scenarios following the metaphor of a virtual theater (Hayes-Roth *et al.* 1997). Their characters are not directly associated with a specific personality. Instead, they are assigned a role and have to express a personality that is in agreement with this role. A key concept of their approach is improvisation. That is, characters spontaneously and cooperatively work out the details of a story at performance time, taking into account the constraints of directions either coming from the system or a human user.

The benefit of agent teams has also been recognized by developers of tutoring systems. For instance, Rickel and Johnson (1999) extended their one-on-one learning environment with additional virtual humans that may serve as instructors or substitute missing team members. The main difference between their work and Avatar Arena is that their agents address the user directly while in our case information is conveyed implicitly by means of a simulated dialogue between the characters.

More recently, Traum and Rickel (2002) have addressed the issue of multi-party dialogues in immersive virtual environments. In the context of a military mission rehearsal application they address dialogue management comprising human-character and character-character dialogues. They propose a layered model for structuring multi-party, multi-conversation dialogues and point out the importance of non-verbal communication acts for turn management. Since the primary field of application of their work is a military mission rehearsal scenario, turn-taking behavior is often predetermined by the distinct roles of the dialogue partners. This is not the case for unmediated negotiation dialogues between equals. Furthermore, besides the “how” to indicate initiative, it is also important to understand the “why” and the “when” a character should try to take the turn in a negotiation dialogue.

Finally, the computer games industry seems to have a rapidly increasing interest in the deployment of multiple characters that interact with the player(s) or

among themselves, e.g., in simulation games, such as the Sims. Unfortunately, technical details concerning the underlying mind models are often not published or even kept confidential.

3. Avatar Arena Version 1

For the purpose of illustration, a meeting date negotiation scenario has been chosen to demonstrate the Avatar Arena framework. This choice is partly motivated by the fact that making a meeting arrangement with others is one of those everyday problems that is often delegated to personal secretaries or assistants. In our scenario we provide a group of users with the ability to have an avatar negotiate with the avatars of other users.

Building a basic version of the Avatar Arena (let us call it Avatar Arena Version 1, or AA-V.1 for short) raises at least the following two modeling tasks: Firstly, the avatars must have some basic understanding of the domain, e.g., they should know what a meeting date is, and that usually one cannot participate in several meetings simultaneously. Secondly, they need some basic conversational skills that will enable them to participate in negotiation dialogues in a way so that a commonly agreed arrangement can be achieved.

3.1 Providing the avatars with basic domain knowledge

An essential part of the domain modeling in the meeting appointment domain is an ontology of meeting dates. Such ontologies have been set up in quite a number of other projects, e.g., for the Verbmobil system that aims at a simultaneous translation of spoken utterances in meeting arrangement dialogues (Wahlster 2000). However, since the primary interest of our research is not on modeling meeting appointment domains, only a rudimentary domain model will be considered here. This model includes a rough characterization of meeting dates along five dimensions: *date*, *type of activity*, *type of person(s) to be met* (family / friends / business partners), *type of temporal fixation* (fixed / movable), and *location* of a meeting.

The emerging taxonomy of meeting dates is shown in the left part of Figure 2. The node “no activity” has been added for the case of an as-yet free slot in an avatar’s calendar. In order to allow a user to enter already scheduled meeting dates, the client-side user interface of the Avatar Arena comprises a calendar-style input widget (cf. the right-hand side screenshot of Figure 2). The entries are then “read” by the user’s avatar and will be taken into account when a new appointment has to be arranged.



Figure 2. Left: Taxonomy of meeting dates. Right: Calendar interface of Avatar Arena

Note that Avatar Arena is designed as a multi-user system. Each user can make “private” settings but neither the user nor her/his avatar has direct access to the settings made by other users. Of course, during a negotiation dialogue, some of the settings may be mentioned in the negotiation process and thus become public.

To a large extent, the contents of the calendar determine what an avatar can say about its availability for a proposed meeting date, and vice versa – which meeting dates the avatar may propose itself in a negotiation process.

For the sake of simplicity, avatars of AA-V.1 have very limited capabilities to reason about time and to solve complex scheduling tasks. They can only check whether or not they (i.e., their users) would be available at a proposed date. If available, an avatar may accept a proposal. If not available, it will reject the proposal and rather propose an alternative date that suits its own calendar.

Also, and again for the sake of simplicity, AA-V.1 avatars do not have a fine-grained system of goals relating to meeting appointments. When entering a negotiation scenario, all avatars have the two goals:

- a. to solve the given task, i.e., to fix the date and location for a meeting;
- b. to preserve their own interests as far as possible.

3.2 Equipping the avatars with basic conversational skills

Meeting appointment scenarios have attracted both researchers studying human-human dialogues (e.g., Rose *et al.* 1995; Alexandersson *et al.* 1997) as well as researchers from the multi-agent systems community who aim at the development of communication protocols for their agents. For example, Shapiro and colleagues (2002) illustrate their proposal for a formal agent specification language (called CASL) by means of a meeting scheduler system. In this multi-agent system, a meeting organizer agent tries to schedule meetings with personal agents which manage the schedules of their human owners. At first glance, this scenario bears great simi-

larity with the application scenario adopted for Avatar Arena. However, there are fundamental differences between the two systems. Firstly, the meeting scheduler deals with disembodied software agents of relatively low functional complexity. Secondly, agents in the meeting scheduler systems use a formalized protocol and a few formalized locutions for symbolic message exchange between machines. In contrast, avatars in Avatar Arena are to represent some human-like qualities that are of relevance in multimodal human-human communication. Therefore, we are more interested in an emulation of dialogues with human participants.

The underlying building blocks of meeting date negotiation processes as discussed here are dialogue moves. All avatars have a number of different communicative acts at their disposal. For AA-V.1 we start with a small repository of acts and group them into different classes which correspond to the different phases of a negotiation:

Opening phase:

- Greet, Reply-to-Greeting;
- Announce necessity to make an arrangement.

Negotiation phase:

- Request-proposal, Propose-date, Accept, Reject.

Closing phase:

- Wrap-up, Leave-taking.

Given the set of possible communicative acts, we need to formulate rules or strategies that can be used for simulating negotiation dialogues.

3.3 Simulation of negotiation dialogues

As pointed out by André and Rist (2001), there are two fundamentally different approaches for the generation of multi-character dialogues. One approach is similar to writing a theater play or movie script for multiple actors. That is, a script author is in full control of all agents in the sense that the author decides on what the single characters are going to utter, how they will react to utterances and actions by others, when to take the initiative etc. Consequently, knowledge about how characters negotiate and converse need to be specified from the point of view of a script-writer (rather than from a character-centric perspective). An advantage of this approach is that it is computationally less complex. In fact, dialogue management can be kept simple because the script author determines all moves in the occurring dialogue games.

In contrast, one can adopt a character-centric approach in which all characters act on their own. That is, each character has to sense the world, recognize

events and actions / utterances made by other characters, and decide on how to react or act appropriately. From the point of view of knowledge engineering, a character's strategic negotiation knowledge and also its conversational skills need to be defined from a character-centric perspective. For instance, each character might have a strategy to prompt a greeting or to reject a proposal that does not fit its own time planning. The attractiveness of this modeling approach lies in its generality as from the point of view of an agent, (a) it might make no difference whether a conversational partner is another synthetic agent or a real human, and (b) the number of negotiation partners can be kept variable and even may change during a negotiation. On the other hand, this generality comes at a high computational cost and demands a dialogue manager that can cope with multi-threaded, multi-party conversations. Also, similar to unmediated meetings, negotiation dialogues that emerge from self-controlled agents can easily become unstructured – if not incoherent – unless the agents are equipped with a broad repertoire of sophisticated conversational skills that enable them to “behave well” but at the same time safeguard their interests in a not necessarily cooperative group discussion.

For Avatar Arena we make a step into the direction of a character-centered system by formulating dialogue strategies from a character-centered perspective. On the other hand, some central control structures are kept in order to avoid various synchronization issues that would arise in a fully distributed system.

Figure 3 shows an excerpt of a meeting appointment dialogue that is chaired by avatar A1. It announces the task of the negotiation and switches to the negotiation phase by making a first proposal. Note that so far we do not distinguish further between direct and indirect speech acts in Avatar Arena. Rather, indirect speech acts are just regarded as linguistic variants of more direct speech acts. Since it is a chaired meeting, all characters will speak one after the other. Given their limited knowledge about solving meeting appointment tasks (cf. Section 3.1), all they can do is to indicate whether or not they are available at a certain proposed date. The negotiation process consists of several circles and terminates if a date is proposed at which all negotiation partners are available.

The lack of temporal reasoning capabilities may cause the avatars appear less competent to a human observer of the generated dialogues. However, reasoning capabilities in the domain of meeting appointment scenarios are part of many commercial calendar systems and could be added to the avatars, too. On the other hand, solely rational approach to appointment scheduling tasks would miss the point of our underlying research objective. In fact, dialogues generated by AA-V.1 have a strong flavor of being nothing more than a mere trace log-file of a multi-agent expert system.

[A1] We have to make a fixed arrangement for a working meeting in the next 8 days.
 [A1] I'm available the day after tomorrow.
 [A2] I'm available then.
 [A3] I'm not available then.
 [A1] I'm available in 6 days.
 [A2] I'm not available then.
 [A1] I'm available in 8 days.

 [A1] Ok, we've done it. We will meet in 5 days.

Figure 3. Excerpt of a sample negotiation dialogue generated by Avatar Arena Version

I N T E R E S T S		Career Hobby Wellness Social Contacts Culture				
		low	medium	high		
	Career	☐	☐	☒		
	Hobby	☒	☐	☐		
	Wellness	☐	☐	☒		
	Social Contacts	☐	☒	☐		
	Culture, Art and Education	☒	☐	☐		
	Work meeting	+				
	Seminar	+				
	Conference	+				
	Work	+				
	Holiday			+	+	
	Med. treatment			+		
	Public event				+	+
	Sport		+	+		
	Hobby		+			
	Family celebr.				+	
	Private meeting				+	

Figure 4. Left: Interface for indicating general attitudes. Right: Relating appointment types to general attitudes

4. Avatar Arena Version 2: Attitudes towards meeting activities

As a next step towards more human-style negotiation dialogues, we introduce the concept of attitudes. That is, we allow the users to assign importance values to already scheduled meeting dates as well as to new dates to be negotiated. We assume that this attribution will vary from one user to another so that we need to assess for each user the importance of all her/his certain meeting dates.

There are several options for making such an assessment. Firstly, we could allow the user to explicitly specify an importance value for each already scheduled meeting. A more fundamental approach is to identify a number of general interests / attitudes to which meeting dates can be related. For Avatar Arena we have taken such an approach and use the attitude dimensions shown in Figure 4. With the form shown on the right-hand side, a user indicates general attitudes. The settings will be used to assess the importance of a certain meeting date for that user. The table on the right-hand side of Figure 4 shows how the assessment is made. A “+” sign indicates a relation between a meeting activity and an attitude dimension. For instance, expressing a high interest in the dimension “career” would assign high importance values to activities that are related to work, such as

participating in working meetings and business events. However, a private activity may also become more important if the activity is carried out together with workmates or the boss.

In AA-V.2 a user specifies already scheduled appointments using the calendar interface (cf. the right-hand side of Figure 4) as well as attitudes towards some value or interest dimensions using the interface shown in Figure 4. To consider the user's attitudes in the negotiation process, we extend the capabilities of our avatars along three directions.

Firstly, we give avatars the freedom to reschedule appointments. An important criterion for an avatar determining its willingness to reschedule an already fixed appointment relates to a comparison of importance values. That is, when a date for the new appointment is proposed, the avatar compares the importance value assigned to the new meeting activity with the importance value of the already scheduled activity. For instance, if an avatar believes in career and does not care much about social contacts, it may be willing to postpone a holiday trip with friends and attend a conference instead. However, there are a number of other factors that an avatar may take into account. In AA-V.2 the avatars also have a notion of time pressure. That is, the more negotiation rounds are needed, the higher becomes the willingness of the avatars to make compromises.

Secondly, we extend the avatar's repertoire of speech acts by including the acts "justify-proposal", and "justify-reject". For instance, an avatar can now argue that it won't be willing to reschedule a medical treatment in favor of a working meeting due to its attitudes. In addition, the importance value assigned to an already scheduled meeting date can be used as a criterion for deciding whether or not to provide a justification at all.

From a technical point of view, we implement the two extensions by formulating additional dialogue strategies for the avatars. Due to these extensions, AA-V.2 negotiation dialogues are somewhat more interesting to listen to as compared to dialogues generated by AA-V.1. Figure 5 shows an excerpt of a typical AA-V.2 dialogue. The effects of the extensions are printed in bold-face.

5. Avatar Arena Version 3: Introducing attitudes towards other Avatars

Dialogues generated by AA-V.2 neglect the fact that in a group setting, single dialogue partners also have attitudes towards their counterparts. Moreover, attitudes towards subject matter, such as meeting activities, as well as attitudes towards dialogue partners, are subject to change in a negotiation process. In functional terms we aim at a simulation test-bed in which a user can also specify her/his attitudes towards other persons (assumed to be represented by their avatars in a

- [A1] We have to make a fixed arrangement for a working meeting in the next 8 days.
- [A1] I'm available the day after tomorrow.
- [A2] I'm available then.
- [A3] I'm not available then. **I have a meeting with my family the day after tomorrow.**
- [A1] I'm available in 6 days.
- [A4] **Well, I wanted to go to the cinema. I may be willing to postpone this.**
- [A2] I'm not available then. **I have another business meeting that day.**
- [A1] I'm available in 8 days.
-
- [A1] Ok, we've done it. We will meet in 5 days.

Figure 5. Excerpt of a sample negotiation dialogue generated by Avatar Arena Version 2

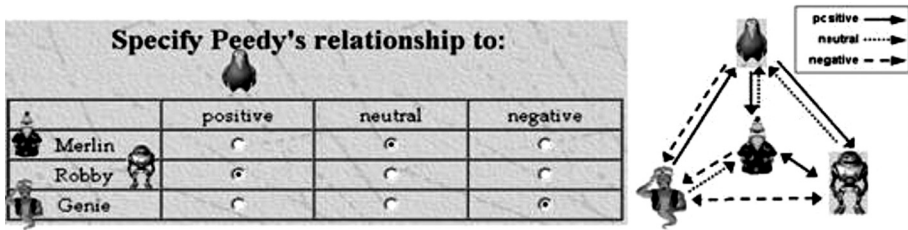


Figure 6. Left: Specifying liking relations. Right: Resulting sociogram

negotiation process). To this end, the user interface of AA-V.3 provides an additional widget that allows the user to specify *a priori* attitudes (or social distances) towards other negotiation partners. For the sake of simplicity, we only distinguish between three different values (positive / negative / neutral) for a liking relationship for each pair of avatars (cf. left-hand part of Figure 6). In the depicted example, the user has chosen Peedy as an avatar, and indicates liking-relationships with regard to the negotiation partners represented by the agents Robby, Genie, and Merlin respectively. That is, Peedy likes Robby, dislikes Genie, and has a neutral attitude towards Merlin.

Interestingly, liking relationships are not necessarily symmetric. That is, a user may like another user (and thus specify that her avatar likes another user's avatar) while the other user actually dislikes her and thus provides her avatar with an opposite setting. Following Moreno (1974) the total of all liking relations as specified by the different users can be used to construct a sociogram (cf. right part of Figure 6).

In our search for an appropriate modeling framework that would capture these aspects we found socio-physiological theories of cognitive consistency,

such as Heider's Balance Theory (Heider 1958), Festinger's Theory of cognitive dissonance (Festinger 1957), and especially "Congruity Theory" by Osgood and Tannenbaum (1955) quite inspiring and also useful with regards to a possible implementation for a further version of Avatar Arena.

5.1 Theories of cognitive consistency

Roughly speaking, Heider's Balance Theory (Heider 1946, 1958) and related derivatives start from the basic hypothesis that a good deal of inter-personal behavior and "social perception" (i.e., the anticipation of attitude assessments) is determined or at least co-determined by simple cognitive configurations which are either balanced or unbalanced. Together with the hypothesis that people tend to avoid unbalanced configurations or cognitive dissonances, one can make predictions about how a certain person may behave in certain social situations. Balance Theory describes the phenomenological world of a person. It establishes a relationship between the person P perceiving the world, another person O, and an object X. To further characterize this situation, the notion of a *cognitive configuration* is introduced, which is in turn characterized by means of two distinct relations:

1. *L-relations* (liking / disliking) to express a person's attitude towards other persons or impersonal entities, e.g., P likes O, P likes coffee, etc.
2. *U-relations* (unit formation) to express a person's perceptive formation of cognitive units, e.g., P perceives the objects X1 and X2 being similar.

Cognitive configurations are always described from the perspective of the person P. They can be balanced or unbalanced. For instance, if P likes O, and P perceives the objects X1 and X2 being similar and at the same time believes that O also perceives X1 and X2 being similar, P's perception of the situation (i.e. P's cognitive configuration) is balanced. Forming combinations of possible relations between P, O, and X yields four kinds of balanced configurations as well as four kinds of unbalanced configurations.

Applying the Balance concept to Avatar Arena we may consider cognitive configurations as illustrated in Figure 7. The left hand diagram of Figure 7 shows a balanced configuration of the parrot Peedy. Peedy likes Merlin but dislikes Genie and also believes that Merlin dislikes Genie, too. In contrast, the right-hand diagram shows an unbalanced configuration. Suppose career is an important value for Peedy who likes Merlin. Up to now, Peedy believed that career is also important for Merlin. However, when talking to Merlin about career opportunities, it turned out that Merlin has no interest at all in issues that relate to career development. In terms of Balance Theory, learning about Merlin's attitude towards the career dimension causes imbalance in Peedy's cognitive configuration.

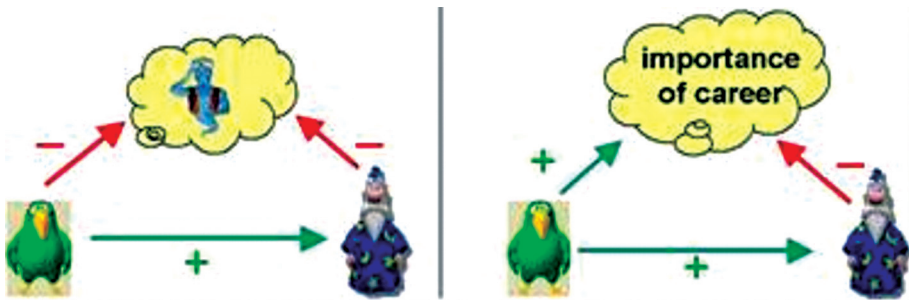


Figure 7. Balanced (left) versus unbalanced (right) cognitive configuration of Peedy

A basic assumption in Heider's theory is that there is a tendency in humans to achieve balanced configurations. As a consequence, a person P with an unbalanced cognitive configuration may perform some cognitive reorganization to achieve balance again. In principle, cognitive reorganization can result in changing:

- P's L-relation to O,
- P's L-relation to X,
- O's L-relation to X, i.e., P makes O change O's L-relation to X,
- several L-relations towards a balanced state.

If due to certain circumstances a balanced configuration is not achievable, the remaining imbalance is likely to produce tension between P and O.

Related to Heider's Balance concept is Festinger's Theory of Cognitive Dissonance (Festinger 1957). Unbalanced cognitive configurations are viewed as cognitive inconsistencies. However, the concept of cognitive consistency / inconsistency is not restricted to Heider's triangular frame of analysis formed by the three entities P, O, and X. Festinger asserts a general tendency for individuals to seek consistency among their beliefs and opinions. In the case of an inconsistency between attitudes or behaviors, there is a tendency to eliminate the dissonance. In particular, a discrepancy between an attitude and behavior may cause a change in attitude so that the attitude accommodates the behavior. Unlike Heider, Festinger also identifies two factors that affect the strength of a dissonance. Firstly, the degree of dissonance depends on the number of dissonant beliefs. Secondly, the degree of dissonance depends on the importance attached to each belief. As a consequence, coping with dissonances becomes a matter of decreasing their strengths. A person may try different strategies for cognitive reorganization to reduce and eliminate dissonance: (a) change the dissonant beliefs to achieve consistency; (b) lower the importance value of dissonant beliefs; (c) increase the number of consonant beliefs to outweigh dissonant ones.

5.2 Linking cognitive consistency to communication

While Heider's and Festinger's work form the foundations of our approach to modeling changes of interpersonal relationships during a negotiation process, we need to combine concepts, such as balance and dissonance with a model of interpersonal communication. Fortunately, attempts in this direction have already been made by some psychologists. For the purpose of building the Avatar Arena the so-called "Congruity Theory" of Osgood and Tannenbaum (1955) deserves special mention. Firstly, Congruity Theory aims at making assertions about attitude changes by means of communication. Secondly, the theory augments the original balance concept in so far that it considers not only polarization but also intensity of (liking) relations.

Similar to the triangle formed by P, O, and X that Heider used in his analysis, Congruity Theory also uses a triangular scheme. This time, however, formed by the triple S, R, X with S being a sender of a message, R being the receiver of a message, and X being the subject matter attributed by the message. Assertions in this model are made from the perspective of the message receiver R. That is, Congruity Theory makes an assertion about the impact of a received message on R. Thereby impact depends on (a) the liking relationship between R and S (from the point of view of R), (b) R's attitude towards the subject matter, and (c) the expressed attitude of S towards the subject matter X.

Similarly to Heider's model, one can distinguish between balanced and unbalanced configurations. For instance, if R likes S, and R has a positive attitude towards X, then a message by S that says something positive about X will result in a balanced configuration formed by the elements S, R, and X. In contrast, if S says something negative about X the resulting configuration S, R, and X would be unbalanced.

Since Osgood and Tannenbaum introduced intensity values to represent the degree of liking or disliking, the distinction between balanced and unbalanced configurations is too crude. Therefore, they further distinguish between balanced, unbalanced and congruent configurations. A congruent configuration is a balanced configuration that in addition requires a match of intensity values. To reduce the combinatorial complexity of their model, however, Osgood and Tannenbaum make the following simplifications:

- there is only a distinction between positive and negative messages,
- the intensity of the liking relation between R and S can take on whole-numbers between -3 and plus $+3$ to express the degree of disliking / liking,
- the intensity of R's attitude towards the subject matter X can vary between the -3 and $+3$ to express the degree of disliking / liking.

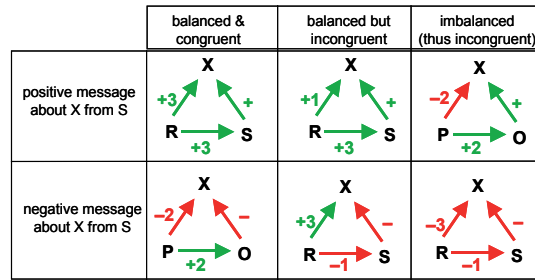


Figure 8. Samples of congruent, balanced, and unbalanced configurations in a receiver's mind after receiving a positive/negative message from S on the subject matter X

Figure 8 presents examples of congruent, balanced and unbalanced configurations after R's reception of either a positive or negative message from S.

If R finds herself in an incongruent or even unbalanced configuration, Osgood and Tannenbaum would agree with Heider and Festinger that something must change. That is, if a message received by R causes incongruence or even imbalance, R will have the tendency to achieve balance and congruency again. Moreover, Osgood and Tannenbaum assert that the magnitude of attitude changes is reciprocally proportional to the original intensities of the relations. This central assertion of Congruity Theory provides the basis for calculating changes in configurations.

For an operationalization of the theory, we adopt the calculus presented in (Herkner 1991, p. 263), as shown in Table 1. Let: I_S be the degree of R liking S, $|I_S|$ the absolute value of I_S , DI_S the delta by which I_S will be changed, I_X the degree of R liking the subject matter X, $|I_X|$ the absolute value of I_X , and DI_X the delta by which I_X will be changed. Using the parameters, the formulae on the right provide the means to calculate changes of intensity values of R's configuration after having received either a positive or a negative message from S concerning X.

Table 1. Operationalization of Congruity Theory

intensity changes for R in response to a positive message of S concerning X	intensity changes for R in response to a negative message of S concerning X
$DI_S = \frac{ I_X }{ I_S + I_X } (I_X - I_S)$	$DI_S = \frac{ I_X }{ I_S + I_X } (-I_S - I_X)$
$DI_X = \frac{ I_S }{ I_S + I_X } (I_S - I_X)$	$DI_X = \frac{ I_X }{ I_S + I_X } (-I_X - I_S)$

5.3 Simulating aspects of group dynamics in negotiation dialogues

Having introduced the concept of cognitive configurations, one may wonder how far this increased “cognitive complexity” of the avatars will lead to a noticeable difference in the negotiation dialogues. With regards to AA-V.3 our working hypotheses are:

1. Before starting a negotiating dialogue, all participating parties have a certain social distance to each other.
2. Avatars make assumptions about the attitudes of other avatars in such a way that the corresponding cognitive configurations are balanced and congruent.
3. When an avatar discovers a mismatch between its assumption about another avatar’s attitudes (i.e., by listening to the other avatar’s utterances), this discovery may cause the experience of dissonance and may eventually trigger a change in the social distance to the other avatar so that a congruent balanced configuration is achieved again.

When considering the specified L-relationships as a rough model of the social context, AA-V.3 can generate much richer negotiation dialogues with respect to:

- negotiation behavior (which goals to pursue, which strategies to apply),
- turn-taking behavior (when to take a turn, when to let others go ahead),
- content selection (what to say),
- linguistic style (how to say),
- non-verbal display (mimics, body postures etc.).

Unfortunately it is not possible within a single project to study in depth the various dependencies between social context and the listed aspects of multi-party dialogues. So far our focus has been on (i) modeling turn-taking behavior for the avatars in an unmediated negotiation setting, and (ii) to enrich the dialogues with personal statements.

In the domain of meeting appointment dialogues, we need to distinguish between several cases in which a character may be motivated to take the turn because it wants:

- a. to make a new proposal for a meeting date,
- b. to reject a proposal because of a conflict with another appointment,
- c. to support a proposal because it fits well with the character’s time planning,
- d. to receive a justification for a proposal from another avatar,
- e. to give a personal comment on a justification of another dialogue partner,
- f. to reply to a personal comment made by another avatar.

[A1] We have to make a fixed arrangement for a working meeting in the next 8 days.

[A1] I'm available the day after tomorrow.

[A2] I'm available then.

[A3] I'm not available then. I have a meeting with my family the day after tomorrow.

[A2] **Can't you postpone this? Our meeting is perhaps somewhat more important.**

[A3] **I don't think so. My family is more important than work.**

[A1] **I'm quite pleased to hear that from you.**

[A1] I'm available in 6 days.

[A4] Well, I wanted to go to the cinema. I may be willing to postpone this.

[A2] I'm not available then. I have another business meeting that day.

[A1] I'm available in 8 days.

[A2] I'm available then.

[A4] I'm available then, too

[A3] I'm not available then. I'm doing sports that day.

[A4] **Can't you postpone this?**

[A3] **I'm sorry but sports are very important for me.**

[A4] **I'm disappointed to hear that.**

[A3] **There are more important things than work.**

[A4] **Maybe, but sports? I hadn't expected that from you.**

[A2] **It's always the same with him.**

[A2] I propose the meeting takes place in 5 days.

.....

[A1] Ok, we've done it. We will meet in 5 days.

Figure 10. Excerpt of a sample negotiation dialogue generated by Avatar Arena Version 2

In sum, turn-taking behavior of the Arena characters is steered by a number of factors including the directedness of the dialogue turn, the urgency to react, and the importance assigned to avoiding uncomfortable states, either because of conflicting goals or because of experiencing cognitive imbalance in a social context. Of special interest are the above listed cases (e) and (f) since the link between the concept of cognitive configurations as introduced above and the importance of acting becomes obvious when recalling the fundamental hypothesis underlying all theories of cognitive consistency. That is, unbalanced cognitive configurations are experienced by individuals as uncomfortable and call for changes. Moreover, the stronger the intensity of a dissonance is, the more important becomes the need to act. We conclude this section by presenting an excerpt of a negotiation dialogue that has been generated by AA-V.3 (cf. Figure 10). Again, the enrichments compared to an AA-V.2 dialogue are printed in bold-face.

6. Conclusion and outlook

In this contribution we have introduced Avatar Arena, a test-bed for the simulation of negotiation dialogues among embodied conversational characters. Using the domain meeting appointment negotiations for the purpose of illustration, we showed how the richness of the generated dialogues can be increased by subsequent extensions of the characters' mind models. To this end, we first considered characters that are only knowledgeable (somewhat) in the domain. Unfortunately, the generated negotiation dialogues seemed to share more similarities with records of trace messages from multi-agent expert systems than with human-human dialogues.

In a second version, we also represented a character's attitude towards certain domain concepts (such as meeting activities /dates). This extension of the model was reflected in the generated dialogues by the fact that the characters had a criterion to determine their willingness to make certain compromises (i.e., to postpone an already scheduled appointment), and to provide personal justifications, e.g., when they reject proposals.

The focus of our research, however, aims at a model that can capture some of the group dynamics that are common in human-human negotiation dialogues. To this end, we considered the character's attitudes towards other characters and modeled a character's social context in terms of liking relationships between the character and all other dialogue partners. Our approach was heavily inspired by socio-psychological theories of cognitive consistency originating in Heider's Balance Theory. For the purpose of Avatar Arena, Congruity Theory by Osgood and Tannenbaum appeared quite attractive for modeling the dynamics of liking relations in a dialogue setting. Firstly, it addresses attitude changes due to communicative acts. Secondly, the model is better elaborated than Heider's balance concept. In particular, it introduces intensities for liking relations and it allows prediction of the kind of cognitive reorganization the receiver may perform. Finally, the required calculations for estimating changes of intensity values are fairly easy to implement. When planning dialogue moves, the characters take into account (i) attitudes towards their communication partners, (ii) attitudes to certain subject content (here meeting dates), and (iii) beliefs about another character's attitudes. Moreover, changes in cognitive configurations may be reflected in the dialogue as personal comments on another character's attitude. On the other hand, our current model is a strong simplification of the underlying theories. More recent sociological research, especially "network analysis" (Jansen 1999), suggests that much more complex modeling is required when striving for highly believable virtual personalities.

While the development of the Avatar Arena is work in progress, we already have a series of working demonstrators which differ in the granularity of the underlying models, the repertoires of negotiation and dialogue skills made available to the avatars, and also in the kind of player technology at the front end. Having a test environment that allows us to compare the consequences of different models and negotiation strategies turned out to be advantageous for the exploration and validation of the deployed concepts, and last but not least, for the purpose of demonstrating the work for further discussion. We have chosen meeting date negotiation dialogues as an application domain. However, we do not claim that our demonstrators are contributions to the issue of finding a meeting date more efficiently. Rather, we are only interested in a study of the negotiation process and its variation depending on different character profiles and different social settings. We are aware of the fact that in real life, especially in a business context, people often have to hide their emotions in order to comply with social norms. On the other hand, we do believe that simulation systems of this kind have high potential for educational applications that aim at training people in complex social interactions.

A major gap in our current version is the absence of a qualitative distinction in emotional reactions. While the amount of change in cognitive configurations may be linked to the notion of arousal, and the quality of a change may be related to the notion of appraisal, yet we do not have a fine-grained model for emotion triggering. However, in cooperation with project partners we are currently aiming at an integration of a model for emotion triggering developed by de Rosis and colleagues (2003). This extension will enable us to describe an avatar's affective state qualitatively as well as quantitatively in terms of a set of emotions. In turn, such affective states will be used to enrich the avatars' verbal and non-verbal expressivity (Walker *et al.* 1997; Pelachaud *et al.* 2002).

A further extension of Avatar Arena concerns the ability of the users to specify some personality parameters (Moffat 1997) for the avatars. One approach is to specify profiles along the dimensions of the "Big Five" model as put forward by (McCrae & John 1992). Such personality profiles can be considered when updating an avatar's cognitive configuration, its affective state, and they may also be reflected in an avatar's communication style.

As pointed out earlier, there is yet much room for improvement in the negotiation skills of the avatars. For instance, according to Pruitt (1991) there are four fundamental types of negotiation strategies. It would be interesting to establish a link between these strategy types on the one hand, and an avatar's "mental" state on the other hand.

Another refinement of Avatar Arena concerns the modeling of both expressive speaker and listeners. That is, we would like to model a tight interaction between

a speaker and her listeners during the act of speaking. First steps in this direction have been sketched in (Rist *et al.* 2003). However, much work remains to be done especially with regard to the question of when changes in cognitive configurations take place in the hearer's mind and in turn trigger, for instance, changes in the hearer's facial display and gaze behavior.

Besides some informal collecting of feedback on the demos from colleagues, no profound evaluation work, e.g. on the believability/naturalness of the generated dialogues has been carried out yet. One approach to evaluate believability would be similar to a Turing test in the sense that one would show protocols of negotiation dialogues to human subjects and have them decide on whether they believe that this was a real negotiation between human beings or a dialogue that has been generated by a computer program.

Evaluation work will also help to shed light on the scalability of the approach. From a technical perspective, scalability is a combinatory issue since the introduction of additional negotiation partner increases the complexity of the model exponentially. However, scientifically more interesting is the relation of group size on the one hand, and negotiation and communication behavior on the other hand. In our future work we will investigate what kind of behavioral changes emerge in our avatars when varying the group size.

Work an Avatar Arena has been taken into account in a number of other applications with multiple embodied conversational characters. One of those is an interactive 3D environment in which a dynamic social model is used to simulate plausible navigation behaviours of characters (Rehm *et al.* 2005).

Acknowledgements

The work presented in this paper has been conducted in the EU funded project MagiCster IST-1999-29078. We would like to thank our project partners for comments and suggestions on our approach.

References

- André, E. & Rist, T. (2001). Controlling the Behavior of Animated Presentation Agents in the Interface: Scripting versus Instructing, *AI Magazine*, 22(4), 53–66.
- Alexandersson, J., Buschbeck Wolf, B., Fujinami, T., Maier, E., Reithinger, N., Schmitz, B. & Siegel, M. (1997). Dialogue Acts in VERBMOBIL-2. VerbmobilReport No. 204, DFKI.
- Cassell, J., Pelachaud, C., Badler, N. I., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S. & Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. *Computer Graphics (SIGGRAPH '94 Proceedings)*, 28(4), 413–420.

- De Rosi, F., Pelachaud, C., Poggi, I., De Carolis, N. & Carofiglio, V. (2003). From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent, *International Journal of Human-Computer Studies*, 59(1–2), 81–118.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
- Gratch, J. & Marsella, S. (2001). Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proc. of the Sixth Intl. Conf. on Autonomous Agents* (pp. 278–285). New York, NY: ACM Press.
- Hayes-Roth, B. & van Gent, R. (1997). Story-Making with Improvisational Puppets. In *Proc. of the First Intl. Conf. on Autonomous Agents* (pp. 92–112). New York, NY: ACM Press.
- Heider, F. (1946). Attitudes and Cognitive Organization. *Journal of Psychology*, 21, 107–112.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. NY: Wiley.
- Herkner, W. (1991). *Lehrbuch Sozialpsychologie*, Bern: Verlag Hans Huber.
- Höök, K., Sjölander, M., Ereback, A.-L. & Persson, P. (1999). Dealing with the lurking Lutheran view on interfaces: Evaluation of the Agneta and Frida System. In *Proc. of the i3 Spring Days Workshop on Behavior Planning for Lifelike Characters and Avatars* (pp. 125–136). Sitges, Spain.
- Jansen, D. (1999). *Einführung in die Netzwerkanalyse. Grundlagen, Methoden, Anwendungen*. Opladen: Leske und Budrich.
- McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its implications. *Journal of Personality*, 60, 175–215.
- Moffat, D. (1997). Personality parameters and programs. In R. Trappl & P. Petta (Eds), *Creating personalities for synthetic actors* (pp. 120–165). New York: Springer.
- Moreno, J. L. (1974). *Die Grundlagen der Soziometrie. Wege zur Neuordnung der Gesellschaft*. Opladen: Leske + Budrich.
- Nitta, K., Hasegawa, O., Akiba, T., Kamishima, T., Kurita, T., Hayamizu, S., Itoh, K., Ishizuka, M., Dohi, H. & Okamura, M. (1997). An experimental multimodal disputation system. In *Proc. of the IJCAI '97 Workshop on Intelligent Multimodal Systems* (pp. 23–28).
- Osgood, C. & Tannenbaum, P. (1955). The principle of congruity in the prediction of attitude change. *Psychology Review*, 62, 42–55.
- Paiva, A. (Ed.) (2000). *Affect in Interactions: Towards a new generation of interfaces*. [LNAI series 1814]. Heidelberg & Berlin: Springer-Verlag.
- Pelachaud, C., Carofiglio, V., De Carolis, B., De Rosi, F. & Poggi, I. (2002). Embodied contextual agent in information delivering application. In *Proc. First Intl. Conf. AAMAS* (pp. 758–765). New York, NY: ACM Press.
- Prendinger, H. & Ishizuka, M. (2001). Social Role Awareness in Animated Agents. In *Proc. Fifth Conference on Autonomous Agents* (pp 270–377). New York, NY: ACM Press.
- Pruitt, D. G. (1991). Strategic Choice in Negotiation. In J. W. Breslin & J. Z. Rubin (Eds.), *Negotiation Theory and Practice* (pp. 27–46). Cambridge, MA: PON Books.
- Rehm, M., André, E. & Nischt, M. (2005). *Let's Come Together – Social Navigation Behaviors of Virtual and Real Humans*. In *Proc. of INTETAIN 2005* (pp. 122–131). Berlin, Heidelberg: Springer.
- Rickel, J. & Johnson, W. L. (1999). Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence*, 13, 343–382.
- Rist, T., Schmitt, M., Pelachaud, C. & Bilvi, M. (2003). Towards a Simulation of Conversations with Expressive and Embodied Speakers and Listeners. In *Proc. of Computer Animation and Social Agents (CASA'03)* (pp. 5–10).

- Rose, C. P., Di Eugenio, B., Levin, L. S. & Van Ess-Dykema, C. (1995). Discourse processing of dialogues with multiple threads. In *Proc. of the ACL* (pp. 31–38).
- Shapiro S., Lesperance, Y. & Levesque, H. J. (2002). The Cognitive Agent Specification Language and Verification Environment for Multiagent Systems. In *Proc. First Intl. Conf. AAMAS* (Vol. 1, pp. 19–26). New York, NY: ACM Press.
- Traum, D. & Rickel, J. (2002). Embodied Agents for Multi-party Dialogue in Immersive Virtual Worlds. In *Proc. First Intl. Conf. AAMAS* (Vol. 2, pp. 766–773). New York, NY: ACM Press.
- Wahlster, W. (Ed.) (2000). *Verbmobil: Foundations of Speech-to-Speech Translation*. Heidelberg: Springer-Verlag.
- Walker, M. A., Cahn, J. E. & Wittaker, S. J. (1997). Improvising linguistic style: Social and affective bases for agent personality. In *Proc. First Intl. Conf. on Autonomous Agents* (pp. 270–277). New York, NY: ACM Press.

Semi-autonomous avatars

A new direction for expressive user embodiment

Marco Gillies, Daniel Ballin, Xueni Pan and Neil A. Dodgson

1. Introduction

Computer animated characters are rapidly becoming a regular part of our lives. They are starting to take the place of actors in films and television and are now an integral part of most computer games. Perhaps most interestingly, in on-line games and chat rooms they are representing the user visually in the form of avatars, becoming our on-line identities, our embodiments in a virtual world. Currently on-line environments such as “Second Life” are being taken up by people who would not traditionally have considered playing games before, largely due to a greater emphasis on social interaction. These environments require avatars that are more expressive and that can make on-line social interactions seem more like face-to-face conversations.

Computer animated characters come in many different forms. Film characters require a substantial amount of off-line animator effort to achieve high levels of quality; these techniques are not suitable for real time applications and are not the focus of this chapter. Non-player characters (typically the bad guys) in games use limited artificial intelligence to react autonomously to events in real time. However, avatars are completely controlled by their users, reacting to events solely through user commands. This chapter will discuss the distinction between fully autonomous characters and completely controlled avatars and how the current differentiation may no longer be useful, given that avatar technology may need to include more autonomy to live up to the demands of mass appeal. We will firstly discuss the two categories and present reasons to combine them. We will then describe previous work in this area and finally present our own framework for semi-autonomous avatars.

2. Virtual characters

This work brings together the two areas of research in virtual characters: *avatars*, which are controlled directly by the users, and *autonomous virtual characters*, whose action and behaviour are controlled by artificial intelligence.

Virtual characters that graphically represent a human user in a computer-generated environment are known as “avatars”. This idea of an avatar synonymous with a user’s identity in cyberspace became accepted after the science fiction novel *Snow Crash*, written by Neil Stephenson (1992). The word “avatar” comes from the ancient language of the Vedas and of Hinduism, known as Sanskrit. It traditionally meant a manifestation of a spirit in a visible form, typically as an animal or human. Examples of modern avatars can be found in virtual worlds, online computer games, and chat rooms. A lot of work has gone into developing graphically realistic avatars; this technology is now being refined and is already commercialised. However, as Ballin and Aylett (2000) point out, believable virtual characters are the summation of two key components: visual realism and behaviour. Therefore it should come as no surprise that current research is now equally focusing on behavioural attributes such as the avatar’s gait and body language, and the user’s individual mannerisms as captured and expressed in their avatar.

The second thread of related research has focused on virtual characters that act independently in a virtual world. These are typically referred to as autonomous virtual characters or virtual agents, and their roots stem from the area of artificial intelligence. Unfortunately for new researchers in the field, several names for these embodied entities have appeared: examples include believable and synthetic characters or virtual agents. Autonomous virtual characters have control architectures designed to make the character “do the right thing” and these usually include a sensor-reflect-act cycle. Here the character makes its decisions based on what it can sense from the environment and the task it is performing. This is compared to other virtual character applications where decisions are based on a set of predicted outcomes. This means an autonomous virtual character needs a sensory coupling with its virtual environment. Naturally, just like any autonomous agent (such as a human or dolphin), it is fallible and will make mistakes sometimes: this could be for several reasons, such as when it might base its decision on incomplete information. However in many respects this makes the character more believable, as we do not act like gods or zombies.

The designers of architectures for autonomous animated characters have taken their inspiration from the AI agent community, and they typically fall into one of two camps. At one extreme lie traditional top-down, planner-based, deliberative or symbolic architectures that typically rely on a world model for verifying sensory information and generating actions in the virtual environment. The

information is used by an AI planner to produce the most appropriate sequence of actions. A good example of an autonomous character using a deliberative architecture is that of STEVE (Johnson *et al.* 1998), a virtual tutor who acts as a mentor for trainees in maintenance of gas turbines in US navy ships or the Mission Rehearsal Exercise, a training system for peacekeepers (Rickell *et al.* 2002). Both architectures are based on SOAR (Laird *et al.* 1987), a mature symbolic AI system that makes sure the sequence of actions in the world are followed correctly. At the other end of spectrum lie autonomous control architectures that are bottom-up and come from non-symbolic AI. These are referred to as Behavioural architectures. These are based on tightly coupled mappings between sensors and motor responses; these mappings are often competing, and are managed by a conflict resolution mechanism. It is the many interactions between the sensed signals in the environment and internal drives that produce an overall “emergent” behaviour. Examples of behavioural approaches can be seen in Terzopoulos and Tu’s (1994) fish, Ballin and Aylett’s (2000 2001) ‘Virtual Teletubbies’, or Grand and Cliff’s (1998) ‘Creatures’. In the case of the Virtual Teletubbies, a robot-based architecture was modified to recreate fictional television characters for children’s entertainment, and offer a level of interaction and stimulation that could not be provided by the television programme.

Of particular interest to us are autonomous characters that can interact with people using appropriate non-verbal communication skills (Vinayagamoorthy *et al.* 2006): examples include Gandalf (Thórisson 1998), Rea (Cassell *et al.* 1999) and Greta (Pelachaud & Poggi 2002). Many characters are also programmed with models of human social relationships that enable them to interact appropriately. Examples in this volume include Rist and Schmitt’s chapter, where the characters have a model of their attitude both to other characters and to concrete and abstract objects in the world. This enables them to negotiate with other characters and establish satisfactory relationships. PACEO by Hall and Oram (also this volume) is an autonomous agent that appears to display an understanding of power hierarchies in an office environment and uses this to interact appropriately with real people.

The work we have presented up to now has made a firm distinction between characters that are directly controlled by a human user (avatars and characters in animation packages) and those that are intelligently controlled by a computer (autonomous agents). This seems a logical distinction, and one that has generally divided the research into animated characters along two general directions: those where the character has no intelligence, such as avatar systems or in an animation, and intelligent virtual agents, who have some degree of self-control, such as the next generation of web hosts. The idea that an avatar could have any degree of autonomy had been seen by many researchers as foreign, or even an oxymoron.

However, increasingly researchers are seeing the importance of bridging this divide. Just because an avatar represents a user, does not mean that it has no independence and cannot exhibit some autonomous behaviour. The next section will firstly discuss the motivation for this sort of semi-autonomous character and then describe a number of similar, existing systems. After that we will discuss our own approach to creating semi-autonomous characters and then describe our implementation of autonomous gaze behaviour.

3. Semi-autonomous avatars and characters

People are constantly in motion, making often very subtle gestures, posture shifts and changes of facial expression. We do not consciously notice making many of these movements and neither do we consciously notice others making them. However, they will contribute to our subconscious evaluation of a person. In particular, when an animated character lacks these simple expressive motions, we clearly notice their absence and judge them as lifeless and lacking personality. We would, however, often find it hard to put our finger on what it is exactly that is missing. The behaviour itself is extremely complex and subtle: LaFrance, in this volume, gives an excellent example with her discussion of vast variation and number of meanings that are possible with as seemingly simple an action as a smile. These expressive behaviours are particularly important during conversations and social interactions.

3.1 Avatars and chat environments

Eye gaze and gesture play an important part in regulating the flow of conversation, determining who should speak at a given time, whereas expressive behaviours in general can display a number of intra-personal attitudes (e.g. liking, social status, emotion). These factors mean that this sort of expressive behaviour is very important for user avatars, particularly in social chat environments. Vilhjálmsón and Cassell (1998), however, note that current graphical chat systems are seriously lacking in this sort of behaviour. Interestingly they note that the problem is not that there is no expressive behaviour but that the behaviour is disconnected from the actual conversations that are going on, and so it loses most of its meaning. This is partly due to the limited range of behaviour that is currently available but they argue that the problem is in fact a more fundamental flaw with avatars that are explicitly controlled by the user. They note four main problems with this sort of system:

1. Two modes of control: at any moment the user must choose between either selecting a gesture from a menu or typing in a piece of text for the character to say. This means the subtle connections and synchronisations between speech and gestures are lost.
2. Explicit control of behaviour: the user must consciously choose which gesture to perform at a given moment. As much of our expressive behaviour is subconscious the user will simply not know what the appropriate behaviour to perform at a give time is.
3. Emotional displays: current systems mostly concentrate on displays of emotion, whereas Thórisson and Cassell (1998) have shown that envelope displays¹ – subtle gestures and actions that regulate the flow of a dialog and establish mutual focus and attention – are more important in conversation.
4. User tracking: direct tracking of a user's face or body does not help, as the user resides in a different space from that of the avatar and so features such as direction of gaze will not map over appropriately.

Vilhjálmsón and Cassell's first two points refer to the problems with simple keyboard and mouse style interfaces, while point 4 shows that more sophisticated tracking type interfaces have problems of their own. Point 3 concerns a type of expressive behaviour that is not directly relevant to the discussion on semi-autonomous avatars. The major problem with the keyboard and mouse interface is that it can only input a small amount of information at a time; it is simply not possible to control speech and gesture at the same time using only two hands. Even if it were possible to create a new multimodal input device that could allow simultaneous control of both speech and gesture, it would be too great a cognitive load for the user to be constantly thinking what to do in each modality. Even if this were not so, point 2 makes it clear that we would not know which gestures to select as so many important signals are subconsciously generated. All this suggests that traditional interfaces are too impoverished to directly control an expressive avatar. Vilhjálmsón and Cassell's answer to point 4 is to add autonomous behaviours that control the avatar's expressive behaviour while leaving the user to

1. There is often a distinction made between envelope and emotion in expressive behaviour. We wonder if there is another type of behaviour that is less basic to conversation than envelope behaviour but more important in day-to-day conversation than emotional expressions. This is the sort of behaviour that expresses and influences intra-personal attitudes and relationships. Whereas envelope behaviour controls the low level, moment-by-moment details of the conversation, intra-personal behaviour might control the high-level relationships between the speakers. Examples might be expression of liking or social status. There could also be more short-lived examples such as behaviour that encourages another speaker to express an opinion or behaviour involved in trying to win an argument.

control the avatar's speech. This creates a new type of animated character that sits between the passively controlled avatar and the autonomous agent. In the rest of this section we will develop Vilhjálms-son and Cassell's argument that this sort of semi-autonomous avatar is important for graphical chat type situations and then describe how it can be extended to other domains.

New interfaces that track the user's face and body might seem to offer an answer to this problem. They could track behaviour without the user having to explicitly think about it and could pick up subconscious cues. However, Vilhjálms-son and Cassell's point 4 argues that for desktop systems this is not possible. The position in space of the user sitting at a computer is very different from that of the avatar, and so their actions will have different meanings. For example, the user will generally look only at the computer screen while the avatar should shift its gaze between its different conversational partners. Vilhjálms-son and Cassell suggest that this sort of interface is only suitable for immersive systems.

However, even here there are problems: clearly full body tracking systems are large, expensive, and currently impractical in a domestic setting, but a worse problem is that even these complex systems are rather functionally limited. They only have a limited number of sensors and these can be noisy, thus giving only a partial view. With face tracking this is even more problematic, especially when the data must be mapped onto a graphical face that can be quite different from that of the user. These deficiencies might only introduce small errors but small errors can create a large difference in interpretation in a domain as subtle as human facial expression. There is a final problem with tracking systems; a user might want to project a different persona in the virtual world. Part of the appeal of graphical chat is to have a graphical body very different from our own. The effect of the tough action hero body would be ruined if it had the body language of the bookish suburban student controlling it.

Before leaving the subject of avatars we should briefly discuss a rather different approach suggested by Michael Mateas (1999), that he calls 'subjective avatars'. This work explores the relationship between the avatar and the user. In current narrative computer games the user tends to control a character with a strong personality and with well-defined goals in the game. However, there is little to guide the user in acting appropriately in role. Current methods tend to be crude, forcing the user down one path. Mateas' text based system uses an autonomous model of the character's attitudes to generate subjectively biased textual descriptions of events that makes the user look through the eyes of the character, instead of a more objective description that leaves the user in doubt as to how to interpret events. This is a very powerful idea potentially very important to the application of semi-autonomous avatars in games. The autonomous behaviour and interpre-

tations of events can then give the user a stronger connection with the protagonist of the game.

3.2 Semi-autonomous characters in other domains

The preceding discussion has focused on the domain of avatars for graphical chat, as this has been the field in which many of these ideas have been developed. However, those ideas are applicable to many other domains where the character does not directly represent the user. The animator generally controls animated characters directly for film but having some of the behaviour autonomously generated could greatly speed the process. This could be very useful for television where budgets are tighter than for feature films.

Moreover, computer-controlled characters do not need to be entirely autonomous. In computer games it is currently popular for the player to have allies that can be controlled indirectly through commands or requests, “Halo: Combat Evolved” is a good example of this. Characters like these can also be classed as semi-autonomous. It might also be useful to have characters that are normally autonomous but whose behaviour can occasionally be influenced or controlled by the director of a virtual environment. This might, for instance, give a teacher the opportunity to guide a child’s use of an educational Virtual Environment. Blumberg and Galyean’s (1995) system is of this type.

3.3 Existing systems and applications

The main problem unique to semi-autonomous avatars and characters is how to combine user input with autonomous behaviour to produce appropriate behaviour for the character. This section will discuss current solutions to this problem and applications of semiautonomous avatars and characters. The main focus of this chapter is on semi-autonomous avatars (i.e. characters that directly represent a user); however, many systems described below involve other types of character. Normally the techniques used are applicable to both avatar and non-avatar characters.

There are two main approaches to combining user control with autonomous behaviour. The first is for the user to give very high-level instructions (“walk over to the door and let Jane in”) and for the character to act autonomously to fulfil them. The character is normally also able to act autonomously in the world without instruction. At one extreme this type of character is manifested in graphical agents that act for the user in a virtual world where the user might not even be present. The user issues instructions or establishes a set of preferences and the

agent thereafter acts autonomously to fulfil these instructions. Examples in this volume include Rist and Schmitt and also Hall and Oram. In both cases, characters act autonomously to negotiate meetings for users in an office environment.

The second approach is to leave some aspects of the character's behaviour to be controlled by the user and others to be controlled autonomously. The focus of this article is primarily on the latter, but most current work falls in the former category so we will spend rather more time discussing it. Though most systems fall into one of these two categories, there is a notable exception in Mateas' subjective avatars (Mateas 1999) described above. In that system, the character's behaviour is entirely controlled by the user but the autonomous system attempts to influence the user into acting in character.

Another important aspect of a semi-autonomous character is the type of behaviour that is produced autonomously. Expressive behaviour such as gesture, facial expression or eye gaze has been studied by researchers such as Cassell, Vilhjálmsón and Bickmore (Vilhjálmsón & Cassell 1998; Cassell *et al.* 2001), Poggi and Pelachaud (1996), Fabri, Moore and Hobbs (this volume), Coulson (this volume), and ourselves. However, it could really be any type of behaviour that is produced currently by autonomous agent; path planning and object manipulation are popular examples.

The final factor we will consider in these systems is the method of user input. Keyboard and mouse are of course popular. Users could directly manipulate the character's body with the mouse, or they could manipulate higher-level features using menus, sliders or other GUI elements. Language-based control is also popular, whether via keyboard, or speech-based. This takes two forms. Firstly, graphical chat, as in Vilhjálmsón and Cassell, where the user enters the text to be spoken and the character autonomously generates non-verbal behaviour based on it. The other type is to give the character high-level linguistic commands, which the character then acts on. Finally, the user's face or body could be tracked and this information, rather than being directly mapped onto the character, could be interpreted and used as input to an autonomous behaviour generation system. This approach may be promising but there has been little work on it so far, see (Vinayagamoorthy *et al.* 2004) for an example. Barakonyi and colleagues (2002) extract MPEG-4 facial action parameters by tracking the user's face, these are used as input to an action generator for their character. This information is then used to reproduce the same emotion etc. but the character might not express it in the same way as the user would have.

Based on these categories, current work can be divided into three main types, discussed below. The first two concern high-level control of autonomous characters, while the last has the user and the computer controlling different modalities in an avatar.

Multi-layered control

Blumberg and Galyean (1995) introduced an autonomous character that could be controlled on a number of different levels, from low-level instructions (for example, issuing commands that directly move parts of their body) to very high-level changes to the characters internal state (for example, making the character more hungry). It is a technique that is generally applied to non-avatar characters but may also be applicable to avatars. Multi-layered control architectures have been popular; for example, Caicedo and Thalmann (2000) created a character that could be controlled by issuing instructions or altering its beliefs. An interesting feature of this system is that it contains a measure of how much the character trusts the user, which influences whether it will carry out the user's instructions. Musse and colleagues (1999) have applied a multi-level system to controlling crowds. Paiva, Machado and Prada (2001) combine direct control of an autonomous character with a more reflective level of control which takes users out of the virtual world allowing users to update the internal state of their character. Carmen's Bright IDEAs (Marsella *et al.* 2000) uses high-level control of the character. Interestingly, the user influences the character's internal state but does not do so explicitly, rather they choose one of three thought bubbles which reflect different state changes. This system will be discussed further in the section on inference below.

Linguistic commands

An obvious way of controlling behaviour of avatars and characters is to give them commands in natural language. For example, Badler and colleagues (2000) implemented linguistic control for avatars in a multi-user VE, and for military training scenarios. Also Cavazza and colleagues (1999) used natural language to control the player character in a computer game modeled on id software's "Doom".

Text chat

We have already discussed this example at length. The user's only input is the text that the avatar should say. Appropriate non-verbal communication behaviour is generated autonomously based on this text. In Vilhjálmsón and Cassell's BodyChat (Vilhjálmsón & Cassell 1998) the avatar produces suitable eye gaze and facial animation to regulate the flow of a conversation. In BEAT and Spark, their follow-up systems (Cassell *et al.* 2001; Vilhjálmsón 2005) they analyse text and determine which gestures should be produced at which particular moments in the text. Similarly, the eDrama system analyses text to extract emotional information that is used for animating avatars (Dhaliwal *et al.* 2007). Poggi and Pelachaud (1996) have done similar work for faces. Gillies and Ballin (2004) use off line customisation, real time commands and recognition of emoticons

to control non-verbal behaviour. Similar methods can also be used for voice, rather than text, interaction. Vinayagamoorthy *et al.* (2002), use an autonomous model of gaze that is triggered by speech in a two part conversational setting. Cassell and Vilhjálmsón, in their evaluation work for BodyChat (Cassell & Vilhjálmsón 1999), discovered that users find the character's behaviour more natural when it is animated autonomously as opposed to when they can control its animation. More surprising was the finding that subjects also felt more in control of the semi-autonomous character. This result is probably due to the fact that users feel overwhelmed at having to control the character's non-verbal behaviour whereas in a semi-autonomous system they can concentrate on the content, such as the speech.

3.4 Future developments

In this section we will describe a number of potential research directions for semi-autonomous avatars and characters. As described earlier the central research problem for semi-autonomous avatars as opposed to other types of agent is the integration of autonomous behaviour and user control. The three areas of research above address this in one of the following ways:

Selective autonomy

Multi-user virtual environments are becoming increasingly heterogeneous, with users of different skill levels accessing them through machines with different capabilities and different interaction devices. Therefore, practical semi-autonomous avatar systems should be designed so that each user can select which parts of the avatar's behaviour are generated autonomously and which are directly controlled, making the set of possible avatars a continuum between complete autonomy (for agents in the world) to complete user control. For example, a world might contain non-user agents which are completely autonomous; text based users whose avatars have autonomous expressive behaviour and also largely autonomous navigation behaviour; desktop graphical users whose expressive behaviour is autonomous but whose navigation behaviour is controlled with the mouse, and finally fully immersed and tracked users whose body motion is directly mapped onto the avatar.

Inferring avatar state

In order to generate appropriate non-verbal behaviour for an avatar, it is useful to know certain things about the internal state of the avatar/user; for example, are they happy, do they like the person they are talking to? One approach might be

to use whatever limited input comes from the user to infer what kind of internal state to project, for example, by analysing the text that the user types. This is of course a hard problem and could easily lead to very inappropriate actions due to incorrect inferences. However, it has the potential to greatly improve the experience. Existing systems such as Spark (Vilhjálmsón 2005) or eDrama (Dhaliwal 2007) use analysis of typed text to infer certain conversational or emotional states of the user. Marsella's Carmen's Bright IDEAs (Marsella *et al.* 2000) supports this type of inference in an interesting way. The user is asked to choose an appropriate thought bubble to represent what the character is thinking. These thought bubbles correspond to changes of internal state but do not expose the user directly to the internal workings of the system.

End-user personalisation

Semi-autonomous avatars should reflect what the user wants them to do as closely as possible and yet with minimum input from them. One way of trying to achieve this is to put some of the work of user control off-line by allowing the user to extensively customise the behaviour of the character before they start to use it. Users of graphical chat systems are very keen to personalize their avatar's appearance (Cheng *et al.* 2002), and there is no reason to believe that this would not be true of behaviour as well. This means not only that avatar behaviour should be very customisable but also that the tools for customizing behaviour should be easy to use for non-expert users. This second requirement is difficult as AI behaviour generation systems are complex and not very easy to understand. Our system, described below, takes a few steps in the direction of building such a tool. Gillies (2006) provides a more complete tool for customising avatars. A different approach that is attracting much interest is the development of mark-up languages that can be used to design the behaviour of virtual humans. Ruttkay and colleagues provide one particularly interesting example in this volume. Their GESTYLE language provides four levels of mark up for specifying differences in style of non-verbal communication between virtual characters.

4. A model for semi-autonomous avatars

We propose a model of semi-autonomous avatars and characters in which the user controls different aspects of the behaviour from the autonomous system. Our model ensures that the autonomous behaviour is influenced by the actions the user performs. This is similar to systems where the user types text and the system generates non-verbal behaviour; however, we allow the user to control certain animated actions while leaving the others autonomous. We divide behaviour into

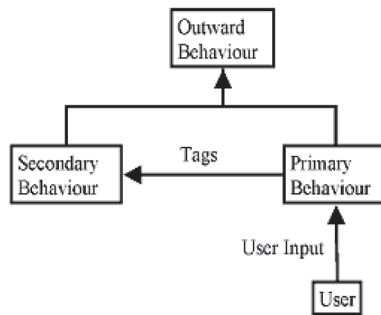


Figure 1. The relationship between primary and secondary behaviours

primary behaviour, which consists of the major actions of the character and is controlled by the user, and *secondary behaviour* that is more peripheral to the action but may be vital to making the avatar seem alive.

For example, a primary behaviour would be invoked if the user requests the avatar to pick up a telephone and to start talking. Secondary behaviour accompanying this might be a head scratch or fiddling with the telephone cord. In our system the primary behaviour can be *tagged* so as to provide a way of synchronising the secondary behaviour. Figure 1 gives an overview of the architecture that is being proposed for primary and secondary behaviour. The primary behaviour is controlled by direct user commands. The secondary behaviour is a module (or set of modules) that is not directly influenced by user input and which acts to a large degree autonomously. To ensure that the secondary behaviour is appropriate to the primary behaviour, it is influenced by messages sent from the primary behaviour module. These messages contain instructions for the secondary behaviour to change appropriately based on the state of the current primary behaviour. Various points in the primary behaviour are assigned tags that result in a message being sent when that point is reached. The tags contain the content of the message. For example, in a conversational system a tag could be attached to the point at which the avatar stops speaking and this could result in various secondary actions being requested from the secondary behaviour module, for example, looking at the conversational partner. The tags should be probabilities of sending a message and the parameters of the message should also be expressed as probabilities. This ensures that behaviour is not entirely deterministic and so does not seem overly repetitive.

There are two ways in which the tags could be edited. The first is when a designer of a virtual environment would want to design the behaviour traits of the characters in their environment. This would be a professional, trained in using the editing package. The end-user would also want to customise the behaviour

of their particular avatar. They, however, would require easy-to-use tools and less ambitious edits. Designers could be given a tool that allows complete control of tags, allowing them to place the primary behaviour tags and edit all of their content. The end-user would be given a tool with more limited control, merely altering certain parameters of the tags, without changing their position. For example, the designer might add a tag requesting that the avatar should look at the conversational partner at the end of an utterance. The end-user might then indicate whether this should be a brief glance with just the avatar's eyes or whether the avatar should orient itself towards the partner with its head and shoulders and look at the partner for a longer time.

4.1 Example: Eye gaze

We have implemented an example of this general architecture for generating eye gaze while an avatar obeys commands given by the user. Eye gaze is a very expressive part of human behaviour and one of the most important cues we use when "reading" other people. This is of course true of gaze between people in social situations such as conversations, giving envelope cues such as that for turn-taking behaviour as well as giving information about social attitudes such as liking. There has been extensive work on simulating this use of gaze, for example (Vilhjálms-son & Cassell 1998; Colburn *et al.* 2000; Vinayagamoorthy *et al.* 2004). However, non-social uses of gaze can also be important in interpreting people's behaviour. What a person is looking at gives a strong indication of their intentions and what they are thinking about. Having a character look at an object before reacting to it makes clear what the reaction was to and so makes the characters behaviour easier to understand. Non-social gaze has been studied by Chopra-Khullar and Badler (1999) but they did not investigate in detail how to integrate simulation of gaze with user control of the avatar's actions. We focus on creating a tool by which a user without programming knowledge can create both primary actions that an avatar can perform as the user requests it, and secondary gaze behaviour that will accompany these primary actions, as summarised in Figure 2.

Our primary behaviour consists of simple actions that an end user can invoke in real time. Each action has one or more targets, which are objects that the character interacts with during this activity. For example, a target for a drinking motion would be a cup. The user would invoke the action by clicking on a possible target. Our aim is to make it easy for the designer of a virtual environment to design a new action. The designer first chooses a piece of motion on which to base the actions and adds some mark-up information. They then designate targets for the action. When the action is invoked the motion is transformed using

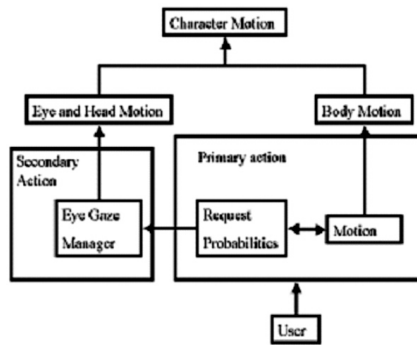


Figure 2. Primary and secondary behaviours for the gaze example

motion-editing techniques (see Gleicher 2001, for an overview) to be appropriate to the new position of the target. For a more detailed description of the primary behaviour see (Gillies 2001).

Secondary behaviour consists of gaze shifts that are controlled by an eye gaze manager described in more detail in (Gillies & Dodgson 2002). The manager can generate eye gaze autonomously and react to events in the environment. The eye gaze can be controlled by sending requests for gaze shifts to the manager, causing the character to look at the target of the request. The gaze behaviour can be controlled by editing one of two types of parameters. Firstly there are parameters that control the character's behaviour as a whole. For example, observing people we noticed that they vary their horizontal angle of gaze but keep their vertical angle relatively constant.² Thus we introduce two parameters to control the character's behaviour: a preferred vertical gaze angle and a probability of maintaining this angle. Setting the parameters in advance allows some end-user customisation of the behaviour. The second type of parameter is attached to a request, changing the way in which the character looks at the target of the request, for example, changing the length of gaze.

As described above, the primary behaviour is tagged with messages that are sent to the secondary behaviour module. In this case the messages consist of eye gaze requests. The designer of the action will add tags to various points in the original motion. These tags will contain a request to gaze at one of the targets of the action, as well as the probability of sending that request. When that point in the motion is reached, the request will be sent with that probability, ensuring that eye gaze can be synchronised with the motion. Values for the parameters of the

2. Though this point is not generally mentioned in the literature it is actually very important. If an avatar's head is made to move vertically too much it looks very wrong.

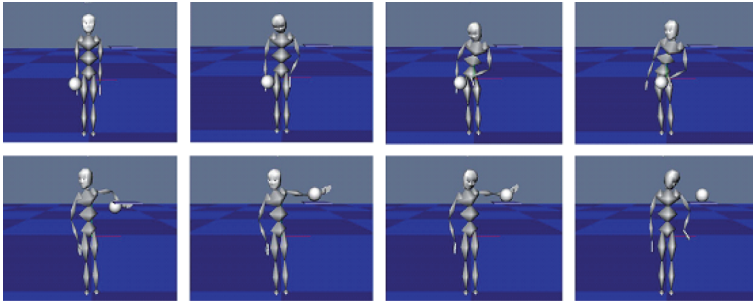


Figure 3. An action of an avatar drinking from a can

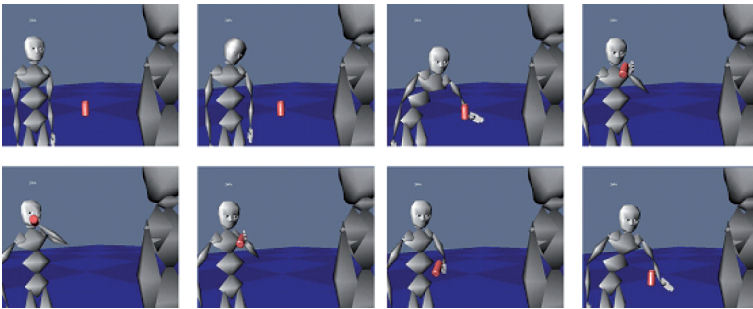


Figure 4. An action of an avatar picking up an object and putting it down somewhere else

request can also be specified, allowing finer control of the gaze behaviour. The designer can also specify what parameters of the tags, including the probabilities, can be edited by the end user. This allows the end user to perform a certain degree of customisation. These parameters are set with a simple interface consisting of a slider for each parameter.

Results and evaluation

Figures 3 and 4 give examples of actions with eye gaze attached. The first is of an avatar drinking from a can. The underlying gaze parameters are set so that the avatar has a tendency not to look around itself and to mostly look downwards when there are no explicit requests. There are two requests tagged to the actions. The avatar looks at the can before picking it up and then at the other avatar shown in the last frame, this time just glancing and moving its eyes without turning its head. This behaviour might indicate avoiding the gaze of the other avatar, which would have a strong intra-personal meaning. The second example is of an action where the avatar picks up an object and puts it down somewhere else. Here the

avatar looks around itself more. There are two tagged gaze requests, to look at the object as it is picked up and at the shelf as it is put down. This time, when the character does not have a request in the middle of the sequence it looks at a location in the distance.

This is a first prototype of this framework, and we are not yet ready to do a formal evaluation. In our opinion the quality of the behaviour is reasonable but could be improved through more careful tagging of the primary behaviour. People viewing the system informally have reported that they consider the addition of eye gaze to add life to the characters and the connection to the primary behaviour gives a stronger sense of intentionality to the character.

Both semi-autonomous avatars in general and our particular system have a large potential for further development. As our system is a general framework, there is a potential to apply it to many different domains and different types of secondary behaviour. There are also specific improvements that could be made to our current implementation. The tool we have described here is still a prototype and needs to be made more robust and tested by creating a wider range of actions and performing user tests. In particular we would like to develop it into a tool that can be used in shared virtual environment and assess people's perception of avatars using our secondary behaviour. As the work focuses on animated actions rather than conversation it would be better suited to a task-based environment than a purely social one. This could form the basis of a formal evaluation of the system. An experiment could be run to compare the user's experience with and without the use of secondary behaviour. The experiment might involve a task that consists of collaboratively manipulating the world using a repertoire of actions.

One aspect that we would like to improve is the user interface for adjusting the various parameters of the secondary behaviour. These allow the user a degree of control over how a particular avatar performs its gaze behaviour. However, these are currently edited using a large set of sliders that directly affect the parameters, some of which are rather counter-intuitive: we would like to provide a more sophisticated and intuitive design tool. Though this model of eye gaze is reasonably general, it is not quite sufficient to model the nuances of interpersonal eye gaze in social situations and we would therefore like to include more heuristics for social situations.

4.2 A conversational character

The framework we have presented is applicable to a number of different uses of characters. This section will briefly describe another application to a character that is able to have a conversation with a real person in an immersive virtual environ-

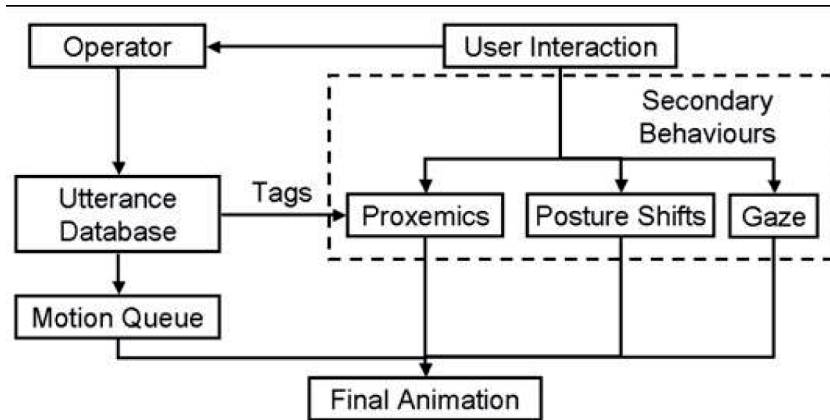


Figure 5. The Architecture for a conversational character

ment. The character is designed for use in virtual reality experiments. The conversation itself is controlled in a “wizard of oz” manner. This application is closely related to the text chat avatars discussed earlier as the character is controlled by a human operator. However, the operator, rather than creating arbitrary textual responses, chooses from a number of pre-recorded audio files of speech responses.

Figure 5 shows the architecture of the character. As in our previous example, the character’s behaviour consists of Primary Behaviour that is triggered by the operator and Secondary Behaviour that occurs largely autonomously in parallel to the Primary Behaviour. In this case, the Primary Behaviour consists of a set of multi-modal utterances that the operator can choose via a graphical user interface, in response to the speech of the user that is interacting with the character. A multi-modal utterance consists of an audio clip containing speech but can also contain other animation elements such as gestures and facial expressions. The secondary behaviour consists of a number of components that respond directly, and in real time, to the behaviour of the user. The user that is interacting with the character has their position tracked and their voice recorded with a microphone. The secondary behaviours can respond in a number of ways to these inputs. The character has three secondary behaviours:

- Proxemics: the character maintains a comfortable conversational distance to the user, stepping forward if the user is too far away or backward if they come too close based on the position tracker.
- Posture Shifts: the character will shift posture occasionally. It will attempt to create a rapport with the user by synchronising its posture shifts with those of the user. This is done by triggering a shift when a large movement is detected from the position tracker.



Figure 6. A conversational character interacting with a human user

- Gaze: the character contains a gaze model based on that of Vinayagamoorthy and colleagues (2004). This model changes the degree of gaze at the user depending on whether the character is talking or listening to the user (as detected by the microphone).

As well as directly responding to the user, the secondary behaviour can also be influenced by the multi-modal utterances selected by the operator. As described in the previous example, the utterances can be tagged with information about the parameters of the secondary behaviours and how they can be changed. For example, a more intimate topic of conversation can be tagged with a closer conversational distance for the Proxemics behaviour. Similarly, any significantly long speech will change the level of gaze at the user in the Gaze behaviour.

This architecture has been used for characters in a number of different experiments (Figure 6 shows an example). The use of Secondary behaviours has proved very helpful in the experimental setting. Firstly, it makes it possible to have a very rich set of behaviour without overloading the operator with excessive work. Secondly, the Secondary Behaviours can respond instantly to the actions of the users without a lag created by the operator's response time. This makes it possible to create responsive effects like synchronization of posture shifts that would be otherwise impossible.

5. Conclusion

We have given an overview of the reasons why semi-autonomous avatars and characters are an important research area, described current research, and suggested possible future directions. We have also presented a framework for semi-autonomous characters, and described an application of this framework to gen-

erating eye gaze. We think this has provided a good demonstration of our general architecture and are pleased with our initial results; however, we are keen to develop these ideas further.

Acknowledgements

Some of this work was done at Cambridge University Computer Laboratory and funded by the UK Engineering and Physical Sciences Research Council. The rest of this work is funded and carried out at BT and at University College London, funded by the UK Engineering and Physical Sciences Research Council. The authors would like to thank the members of the Cambridge University Computer Lab Rainbow research group, the Radical Multimedia Lab, UCL Virtual Environments and Computer Graphics group, Mel Slater and Tony Polichroniadis for their support and suggestions.

References

- Badler, N., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L. & Palmer, M. (2000). Parameterized Action Representation for Virtual Human Agents. In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied Conversational Agents* (pp. 356–284). Cambridge, MA; MIT Press.
- Ballin, D. & Aylett, R. S. (2000). Time for Virtual Teletubbies: The development of Interactive and Autonomous Children's Television Characters. In *Proc. Workshop on Interactive Robotics and Entertainment* (pp. 109–116). Carnegie-Mellon University, April 2000.
- Ballin, D., Aylett, R. S. & Delgado, C. (2001). Towards the development of Life-Like Autonomous Characters for Interactive Media. In *Procs. BCS Conference on Intelligent Agents for Mobile and Virtual Media, National Museum of Film and Photography*. Bradford, UK 2001.
- Barakonyi, I., Chandrasiri, N. P., Descamps, S. & Ishizuka, M. (2002). Communicating Multimodal information on the WWW using a lifelike, animated 3D agent, In H. Prendinger (Ed.), *Proceedings of the PRICAI workshop on Lifelike Animated Agents* (pp. 16–21). Tokyo, Japan, August 19, 2002.
- Blumberg, B. & Galyean, T. (1995). Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. In *Proceedings of ACM SIGGRAPH 1995* (pp. 47–54). ACM Press.
- Caicedo, A. & Thalmann, D. (2000). Virtual Humanoids: Let Them be Autonomous without Losing Control. In *Proc. of the Fourth International Conference on Computer Graphics and Artificial Intelligence*. Limoges, France. May 3–4, 2000.
- Cassell, J., Bickmore, T., Campbell, L., Chang, K. & Vilhjálmsón, H. H. (1999). Embodiment in Conversational Interfaces: Rea. In *Proceedings of ACM SIGCHI 1999* (pp. 520–527). ACM Press.

- Cassell, J. & Vilhjálmsson, H. H. (1999). Fully Embodied Conversational Avatars: Making Communicative Behaviours Autonomous. *Autonomous Agents and Multi-Agent Systems*, 2(1), 45–64.
- Cassell, J., Vilhjálmsson, H. H. & Bickmore, T. (2001). BEAT: The behavior expression animation toolkit. In *Proc. of ACM SIGGRAPH* (pp. 477–486). Los Angeles California: ACM Press.
- Cavazza, M., Bandi, S. & Palmer, I. (1999). Situated AI in Video Games: Integrating NLP, Path Planning and 3D Animation. In *Proceedings of the AAAI Spring Symposium on Computer Games and Artificial Intelligence* [AAAI Technical Report SS-99-02]. Menlo Park, CA: AAAI Press.
- Cheng, L., Farnham, S. & Stone, L. (2002). Lessons Learned: Building and Deploying Virtual Environments. In R. Schroeder (Ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Worlds* (pp. 90–111). Heidelberg & Berlin: Springer.
- Chopra-Khullar, S. & Badler, N. (1999). Where to look? Automating visual attending behaviors of virtual human characters. In *Proceedings of the 3rd Autonomous Agents Conference* (pp. 16–23). ACM Press.
- Dhaliwal, K., Gillies, M., O'Connor, J., Oldroyd, A., Robertson, D. & Zhang, L. (2007). eDrama: Facilitating online role-play using emotionally expressive avatars. In P. Olivier & R. Aylett (Eds.), *Proceedings of the 2007 AISB Symposium on Language, Speech and Gesture for Expressive Characters* April 2–4, 2007, Newcastle University, Newcastle upon Tyne, UK.
- Gleicher, M. (2001). Comparing Constraint-Based Motion Editing Methods. *Graphical Models*, 63, 107–134.
- Gillies, M. (2001). Practical behavioural animation based on vision and attention Cambridge University Computer Laboratory technical report UCAM-CL-TR-522.
- Gillies, M. & Dodgson, N. (2002). Eye Movements and Attention for Behavioural Animation. *Journal of Visualization and Computer Animation*, 13(5), 287–300.
- Gillies, M. & Ballin, D. (2004). Integrating Autonomous Behavior and User Control for Believable Agents. In *Proc. Intl. Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)*. New York, NY: ACM Press.
- Gillies, M. (2006). Applying direct manipulation interfaces to customizing player character behaviour. In R. Harper, M. Rauterberg & M. Combetto (Eds.), *Entertainment Computing – ICEC 2006 – 5th International Conference, Proceedings* (pp. 175–186) [LNAI 4161]. Berlin & Heidelberg: Springer.
- Grand, S. & Cliff, D. (1998). Creatures: Entertainment software agents with artificial life. *Autonomous Agents and Multi-agent Systems*, 1(1), 39–57.
- Laird, J., Newell, E. & Rosenbloom, P. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1), 1–64.
- Marsella, S. C., Johnson, W. L. & LaBore, C. (2000). Interactive Pedagogical Drama. In *Proceedings of the 4th international Conference on Autonomous Agents* (pp. 301–308). ACM Press.
- Mateas, M. (1999). Not Your Grandmother's Game: AI-Based Art and Entertainment. In *Proceedings of the AAAI Spring Symposium on Computer Games and Artificial Intelligence* [AAAI Technical Report SS-99-02]. Menlo Park, CA: AAAI Press.
- Musse, S. R., Garat, F. & Thalmann D. (1999). Guiding and Interacting with Virtual Crowds in Real-Time. In M. Magnenat-Thalmann & D. Thalmann (Eds.), *Proceedings of the Eurographics 1999 workshop on Computer Animation and Simulation* (pp. 23–34). Milan, Italy, September 7–8, 1999.

- Paiva, A., Machado, I. & Prada, R. (2001). The child behind the character. *IEEE Transactions on systems, man and cybernetics: Part A*, 31, 361–368.
- Pelachaud, C. & Poggi, I. (2002). Subtleties of Facial Expressions in Embodied Agents, *Journal of Visualization and Computer Animation*, 13, 301–312.
- Poggi, I. & Pelachaud, C. (1996). Context Sensitive Faces. In R. Campbell & C. Benôit (Eds.), *Proceedings of the ESCA Workshop on Audio-Visual Speech Processing* (pp. 17–20). Rhodes, Greece, 24–27 September, 1996.
- Rickel, J. Marsella, S., Gratch, J., Hill, R., Traum, D. & Swartout, B. (2002). Towards a New Generation of Virtual Humans for Interactive Experiences, *IEEE Intelligent Systems*, July/August 2002, 32–38.
- Stephenson, N. (1992). *Snow Crash*. Bantam Books.
- Terzopolous, D., Tu, X. & Grzeszczuk, R. (1994). Artificial fishes: Autonomous, Locomotion, Perception, Behavior, and Learning in a simulated physical world. *Artificial Life*, 1(4), 327–351.
- Thórisson, K. (1998). Real-time Decision Making in Multimodal Face-To-Face Communication. In *Proc. Second Intl. Conference on Autonomous Agents* (pp. 16–23). New York, NY: ACM Press.
- Vilhjálmsón, H. H. & Cassell, J. (1998). BodyChat: Autonomous Communicative Behaviors in Avatars. In *Proc. Second Intl. Conference on Autonomous Agents* (pp. 269–276). New York, NY: ACM Press.
- Vilhjálmsón, H. (2005). Augmenting Online Conversation through Automated Discourse Tagging. In *Proc. 6th annual minitrack on Persistent Conversation at the 38th Hawaii International Conference on System Sciences*, January 3–6, 2005, Hilton Waikoloa Village, Big Island, Hawaii, IEEE 2005.
- Vinayagamoorthy, V., Garau, M., Steed, A. & Slater, M. (2004). An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *Computer Graphics Forum*, 23(1), 1–11.
- Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C. & Slater, M. (2006). Building Expression into Virtual Characters. In *Eurographics Conference State of the Art Reports* 2006.

The Butterfly effect

Dancing with real and virtual expressive characters

Lizbeth Goodman, Ken Perlin, Brian Duffy,
Katharine A. Brehm, Clilly Castiglia and Joel Kollin

1. Introduction

This paper presents the backstory to the early years of the Butterfly Project, as presented by Lizbeth Goodman and Ken Perlin at the AISB'02 Symposium *Animating Expressive Characters for Social Interactions*, Imperial College, London, 4–5 April, 2002 (<http://homepages.feis.herts.ac.uk/~comqlc/aecsi02>), which introduced the project as a collaboration between the Virtual Interactive Puppetry (VIP) project of the SMARTlab Centre of Central Saint Martins College of Art and Design, the London Institute, and of the CATLAB (the Media Research Laboratory) at New York University. It is also based on the paper given by Brian Duffy at the same event (O'Hare & Duffy 2002). Together, these papers introduced the first phases of what was to become a major international project led by SMARTlab. The research for the Butterfly Project was, in 2002–2003, shared between the three lead investigator's host institutions: the SMARTlab at Central Saint Martins College of Art and Design, London, the CATlab at NYU, and Media Lab Europe, Dublin. These three groups joined forces in 2003 under the new name SPIRITlevel.

This joint research and performance-technology project began life as VIP: Virtual Interactive Puppetry, and evolved into the Flutterfugue described here and then into more recent 2D/3D animation performances. In the years since the findings of the first Butterfly Project were presented, much progress has been made on the project (see e.g., Goodman 2003a, 2003b, 2007a, 2007b; Burke *et al.* 2004) and many institutional changes¹ have also occurred to give the work a

1. The CATlab was closed down in early 2004, but their team of artist-technologists (Castiglia, Kollin, Feeley, Brehm, Sudol, and others) continued to work on a project basis within the SPIRITlevel Consortium. In early 2004, Dr Duffy left Media Lab, first to set up his own robotics group at University College Dublin, then to work at Eurecom and Man-Machine (<http://www.manmachine.org>) in Sophia-Antipolis, France. The SMARTlab moved to the University of East London, its current home: www.smartlab.uk.com/1about/index.htm.

*Ode to a Butterfly –
Stay near me – do not take thy flight!
A little longer stay in sight..... William Wordsworth.*



Figure 1. Lizbeth Goodman and Jaihn K Rose performing with the animated Butterfly controlled by puppeteer Kate Brehm, 2002

different status and grounding. In anticipation of these and other changes, the SMARTlab set up the SPIRITlevel consortium and continues to manage that team in order to support the work of these and other researchers, artists and people with disabilities around the world, as they continue to interact with the Butterfly and all our other 3D creatures and projects in many forms.²

The initial findings presented at the AISB'02 Convention were tested in a live performance/technology collaboration in London in July 2002. This paper includes an evolution of the project and its aim, extracts from the AISB presentations as work in progress, and input from some of the contributing project team who jointly presented the Flutterfugue as part of the Mediatheque event³ in July 2002.

The project is ongoing: this paper summarises its origins, rationale and progress as of the end of 2002, with updates from 2004 and 2007. Full project details and recent research by all the authors can be found on the SMARTlab website – www.smartlab.uk.com.

2. See <http://www.smartlab.uk.com/2projects/index.htm>.

3. Videos of this performance can be seen at www.smartlabcentre.com/mediatheque/flutter.htm.

2. From the chrysalis: The origins of the project⁴

2.1 Virtual Interactive Puppetry

The Virtual Interactive Puppetry (VIP) project began⁵ in 1999, when Lizbeth Goodman and colleagues were experimenting with telematic performance as part of an MA course experiment online, linking dancers and performance specialists around the world, all sharing the common aim of wanting to move with each other, reaching through the computer screen, through time and space. The creation of the shared screen (now known as the SMARTshell) has been documented elsewhere (Goodman 2003a), as have the Extended Body Project, and the VIP project as a whole (Goodman 2003b).

As these initial experiments led to a concrete research project founded on human movement, Goodman realised early on that the most important aspect of the project's success would be the creation of a user-friendly interface and portable technology system. From the outset, the intention was to make the VIP system accessible to dancers and artists with disabilities, and others with limited physical mobility who might wish to regain or construct a new sense of empowerment through virtual movement with puppets and people in real time and space. The system as first envisioned was to be modular, and have many different user outcomes and distribution opportunities. It was, from the beginning, meant to be scaleable, so that it could be used full or in part, depending on the total performance or experiential outcome desired.

The larger VIP Project has been discussed in more depth in previous publications (Goodman 2001) and was the focus of discussion at the Banff Centre in Canada, when Goodman met Perlin and the subset of VIP now known as the 'Butterfly Project'; at that point, collaboration began. This experimental phase of the VIP Project was, in time, to find many audiences with many needs and interests. The current and future iterations of the project are described at the end of this article. But let's begin with the beginning, which in drama and performance is often to do with exploration through movement, costume, mask, voice, gesture. Here we see three phases of development of the Butterfly character as we now know her.

4. By Lizbeth Goodman.

5. The history of the VIP and Butterfly project to date can be found on the VIP CD ROM produced by Mo-Ling Chui for the SMARTlab in 2002. Copies are available upon request from www.smartlabcentre.com.

The VIP Project (Phase One) set out to:

- Construct a new performance environment and platform for virtual interactive puppetry: the CASS system (collaborative augmented stage set).
- Enable collaborative distributed performance over distance in synchronous and asynchronous modes.
- Create frameworks for collaborative awareness in distributed performance spaces linking audiences and performers.
- Apply multi-modal interfaces for capturing body expressions to be used for creating ‘liveness’ in telematic puppets.
- Develop a series of prototype systems enhancing visceral awareness and connections of performers and audiences, informing interaction and performance over distance.

The original planning of the project conceived of a sequence as follows:

1. Puppeteering and playing with ‘input’ puppets generates motion and tactile information.
2. Video-based gesture-capture tracks hand motions.
3. Encoding and mapping onto a virtual actor (‘vactor’).
4. Projection and re-embodiment of ‘vactor’ into a physical, full-size animatronic puppet.
5. Performance with animatronic puppet on smart stage, testing collaborative awareness of participants at each VIP location.

Arising from work over a number of years with graduate students whose bodies could no longer move and dance as they once did, and who were learning to use wheelchairs in dance and movement experiments, Lizbeth Goodman launches some early experiments in telematic performance with people and puppets, choreographed remotely, in order to test levels of control and movement.

She tested this work in the first iteration of the Extended Body experimental MNA course, run in association with her then-institutional base (the Institute for New Media Performance Research at the University of Surrey, UK) and with collaborating students and staff at NYU, ASU, and internationally. This work was, in time, written up and presented the Banff Living Architectures event, including the demonstration and discussion of some new ideas arising from work in progress on a virtual puppetry project developing from that larger area of research. As a contribution to the investigation in a shared space and time (the Living Architectures Conference in Banff in late 2000), this presentation sought to disrupt the notion of a body performing in time and space (Goodman & Kueppers 2001).

In testing of the ideas and discussion with colleagues, Goodman was invited to return to Banff to take the VIP Project into further development. At the same time, Goodman began working as theatre games consultant/producer/dramaturge with Sara Diamond on a separate project of hers: the Code Zebra Project.

2.2 The Code Zebra Project

The Code Zebra Project – for very different reasons to do with the need to disrupt the action dynamics of a contained ecosystem of animal characters with predictable patterns, with the movement of a ‘free radical’ character that could appear and flit through the world at will, distracting, disturbing and generating movement from the other characters – in time evolved a butterfly character as one of the ‘free radicals’ in the visualisation dynamic (along with a snake and a peacock: see www.codezebra.net for details).

In different phases of the work, Goodman played the butterfly and the peacock in the dramaturgical experiments and the live dance club environments. Moving in character as these winged creatures and going through the dramaturgical exercises of ‘getting into character’ using scent, make-up, stretching and flight-related movements such as perching, hovering, etc., all recalled the very first experiments previously undertaken for the VIP project, when Goodman tested the limits of her own physical mobility when restricting her motions to match those of the physical marionettes of the Forkbeard Fantasy show at the Theatre Museum (Figure 2).

The restricted motions of the Forkbeard marionettes were limiting to the performer, who was responding to cues and direction from dancers elsewhere in the world requesting movement in certain ways, while only allowing her body to move as far as the partner marionette could move. However, they were also highly elucidating as a lesson in body architecture and emotional frustration.

Nearly concurrently, another form of movement experiment began for another project, and the two began to dovetail temporarily while the butterfly character was born. So, the VIP puppetry project and the flight of the butterfly for Code Zebra began to develop in parallel (see Figure 3). They soon let go their linked wings and separated again into two distinct projects. But, dancing the butterfly for two audiences, and feeling the restrictions of physical possibility in real and virtual space, Goodman began to feel the need to create a butterfly that could really lift up and fly. So, the butterfly began to dance in live performance experiments, and on screen.



Figure 2. Goodman with Forkbeard puppets

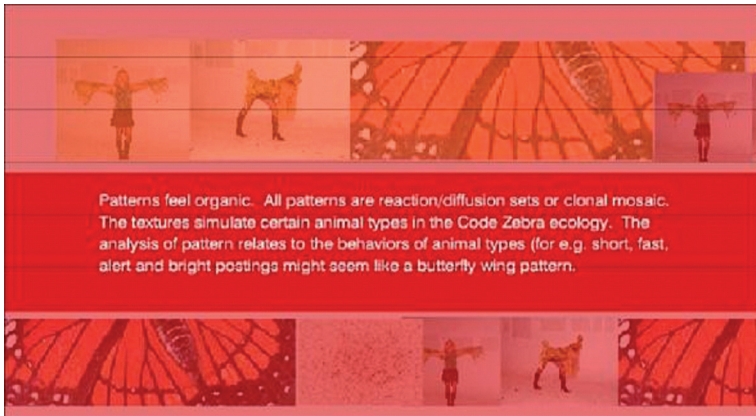


Figure 3. A slide from the Code Zebra Character Development web site and demo CD Rom produced concurrently with our Budapest live performance streaming session, funded by South East Arts (2001). See www.codezebra.net

2.3 Birth of the Butterfly

In the midst of this series of experiments, and while teaching and writing on other issues as well, Goodman returned to Banff to speak again at the Human Generosity Event in August 2001. There, what became known as ‘The Butterfly Project’ was born.

The findings of the work to date were presented and the need to find a flying avatar discussed: something that would provide more freedom of motion yet more equal access to the creation of that motion for movers of all levels of physical

ability. In time, one would need a team to create a series of alternative interfaces that could be triggered by movement, breath or any other body sense. But first, to continue the movement experiments more effectively, the VIP project needed a flying avatar. After giving this talk, discussions with Ken Perlin about the project included the forms that a flying puppet might take.

In the months that followed the Human Generosity event, the idea of making a series of winged creatures, puppets and animations to further the aims of the VIP project was discussed with Ken Perlin, Clilly Castiglia and the team at NYU. A previous MA student of Dr Goodman's, Kate Brehm, was by this time already presenting VIP for the team at major conferences, and was beginning new research for a possible PhD on the VIP project. She had already taken part in the experimental VIP telematic experiments from the London stage to the linked classroom, participating via the SMARTshell from the NYU Drama Department. Kate joined the CATlab at this time to work on the puppetry interfaces on site.

The new Butterfly Project team began to form. This project brought in a new collaboration partner, BBC Imagineering, and focussed in on the role of interactive music and dance with animated and robotic puppetry interfaces when Nick Ryan went with Goodman to New York to further develop a score for our developing winged performance scenario.

3. The creation of the winged puppet⁶

3.1 Artistic rationale

As discussed above, the team found it compelling to consider the flight of an avatar moving with dancers in real space and in shared telematic real-time experiments. As the team worked, they sought to create that 'back of the head' visceral sense of connection and energy that comes from feeling connected to another being at a deep level, real and symbolic.

3.2 Technical implementation

The virtual butterfly was implemented entirely in Java 1.0, so that it would work in Web browsers without requiring any software plug-ins. It is built on top of a software buffer renderer, which supports Phong shading and transparency at real-time

6. Artistic Rationale by Lizbeth Goodman; Technical Implementation by Ken Perlin.

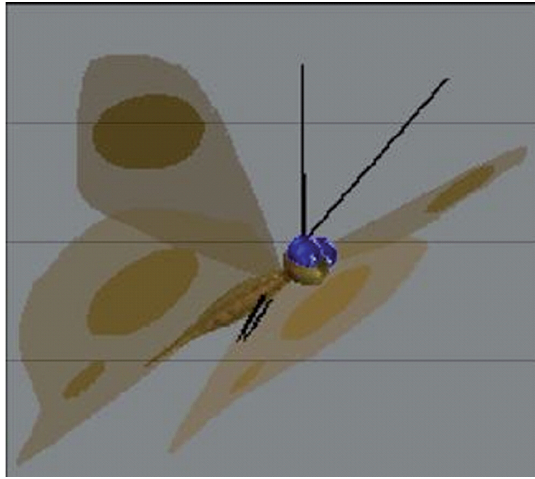


Figure 4. The Butterfly

rates. The underlying modelling and rendering package is accessible on the Web at <http://mrl.nyu.edu/~perlin/render>.

The butterfly (Figure 4) was designed to be a semi-autonomous virtual character. It contains built-in animation routines to handle wing flapping, spin, or “roll” about its longitudinal axis, controlled looping motion, perching behaviour, and pseudo-random fluttering behaviour. Each of these behaviours can be independently modulated by varying the amplitude of a small set of scalar-valued parameters. The behaviours are designed to work with each other; any combination of parameter values produces a reasonable visual result.

At the lowest level, these animation behaviours are all built up from a forward-kinematics jointed model. The left and right wings are independently controllable, each wing consisting of two separate parts: a front part and a rear part. Each wing part can rotate about any axis from its attachment point on the thorax. Each of the high-level behaviours is created as a time-varying set of parameterized signal generators over this low level model. The signal generators modulate component amplitudes in sinusoidal patterns, with varying frequencies and phase offsets, thereby creating the repetitive motions required for flapping, looping, rolling, and so forth.

A puppeteer controls the puppet by varying the amplitudes of the high-level behaviours. The puppeteer does not need to be concerned with the low-level signals that are ultimately generated at the lower level of the individual kinematics joints. This layer of abstraction frees the puppeteer to make high-level aesthetic decisions in the context of a responsive interactive performance. A simple dem-

onstration of the butterfly's behaviours can be found on the Web as a Java applet at: <http://mrl.nyu.edu/~perlin/butterfly>.

3.3 Artistic rationale, part 2

Once a butterfly was up and flying, of course, the team could begin to experiment in all the ways that would be necessary to bring this creature onto the screens and, more importantly, into the lives of the intended audiences. It was agreed to do a first trial as a screen-based optical system to be constructed at the CATlab and then tested against the larger aims.

4. Creation of the Butterfly's environment: The screen and optical systems design⁷

4.1 The Butterfly CG stereo projector (as seen by the audience in a performance context)

For the Butterfly Project, CATlab, working with the SMARTlab designed and built a low-cost, high-precision stereo projection unit for use in a performance environment. Two InFocus 530 DLP-type projectors were hung from special mounting plates, which allow fine 2-axis angle adjustment to allow the images to be superimposed. The mounting plates were mounted in a fixture made from strut material that allowed coarse adjustment for initial set-up. Each projector had a circular polarized filter – one right-handed, one left-handed – providing stereoscopic 3D depth for the viewers (correctly) wearing circularly polarized glasses. Most screens do not preserve polarization and therefore are unsuitable for this type of 3D. A Stewart “Disney Black” screen was used at NYU in both front and rear-projection mode and a Da-Lite silver screen in London.

4.2 Stereo video projector (for use by the puppeteer)

The butterfly's puppeteer can view the performers and butterfly simultaneously through a stereo camera setup which feeds a separate small rear-projection polarized screen. The cameras use left- and right-handed polarizing filters to be compatible with the CG stereo projector described above, and the puppeteer can see

7. By Joel Kollin, creator of the 2D and 3D projection screens.

the Butterfly at the correct depth relative to the performers, as both are scaled down. This system enables remote puppeteering of the Butterfly.

5. The Butterfly Project collaboration phase between SMARTlab and NYU⁸

Ken Perlin's involvement in the Butterfly Project, as outlined above, focused on creating the flying creature as an animation for future development in a range of performance contexts. This short text is included on the CD ROM demonstrator version of the Butterfly Project as well. It outlines the system operation of the first iteration of the Butterfly.

The initial long-term purpose of The Butterfly Project was to develop tools that allow people to work with computer generated objects as though they were real, freely and intuitively defying space and scale. The short-term goal of the Butterfly project was to allow a puppeteer to work with computer graphic avatars, such as a *virtual butterfly*: with much of the same kinaesthetic immediacy and facility with which the puppeteer can work with real puppets. Of course, virtual puppets can do things that it is very difficult to do with real puppets, such as freely change their shape, size and colour, make arbitrary sounds, and appear simultaneously in different places and scales.

The first goal was to enable a creation space in which the puppeteer can perceive, simultaneously with responsive stereo vision and 3D sound, graphically generated puppets that appear to be within the puppeteer's hand/eye space. For this purpose, the puppeteer wears passive stereo glasses (alternately polarized) and wears wireless earphones. Three small LEDs are mounted on the glasses, and two video cameras are trained on the puppeteer from above. The positions of these three LEDs are continually tracked and analyzed by computer, in order to follow the position and orientation of the puppeteer's head. This information is used to vary the computer graphic imagery and 3D sound which the puppeteer perceives, creating the illusion that it is floating in 3D space in front of the puppeteer. It was also intended that the puppeteer could perceive, with stereo vision and 3D sound, a miniature version of a live actor or dancer.

An audience in the theatre looking at the dancers and wearing passive stereo glasses will be able to perceive the dancers interacting with life-size versions of the computer-generated puppets. In one scenario, the puppeteer holds a Polhemus or other 6-DOF sensor in one hand, and controls a set of MIDI sliders, or

8. By Clilly Castiglia, Creative Director of the Collaboration with Ken Perlin.



Figure 5. Actor with Butterfly

some equivalent parametric controller, in the other hand. The hand that holds the sensor literally flies around setting the position and base orientation of the puppet. The other hand varies behaviours. In the case of a butterfly, these behaviours might include rate of wing flap, magnitude of wing flap, spin and looping. The puppeteer hears synthetically generated sound from the puppet from the proper 3D location. In the case of a butterfly, one sound might include wing flapping, which will vary in rhythm, pitch, amplitude and timbre, depending upon how the position and behaviours are set.

The puppeteer sees the live actor in miniature. For this purpose, the actor's performance (see Figure 5) is captured by two video cameras, spaced far apart to create hyperstereo. For example, if the puppeteer is expected to operate at one tenth normal scale (thereby seeing the actor as five to six inches tall), then the two video cameras need to be separated from each other by about ten times interocular distance, or about 25 inches. Because the rear projection in the live performance cannot project the animated puppet so as to be directly in front of the actor (thereby hiding the actor from the audience), it is the puppeteer's responsibility to ensure that during the performance, the puppet never directly upstages the actor. However, the puppet can be made to appear to fly nearer to the audience than the actor, as long as it does so while vertically higher than the top of the actor's head.

A number of new software tools and rehabilitation tools are being developed for this project. For example, a software tool is currently being created to convert the image of the three LEDs already mentioned, as seen by two overhead cameras, into the position/orientation of the puppeteer's head. A set of matched projectors with circularly polarized filters will project the stereo image of the butterfly for the audience.

To give the butterfly's puppeteer a 3D representation of the dancer we place reflective dots on the dancer. Two cameras are positioned in front of the stage,

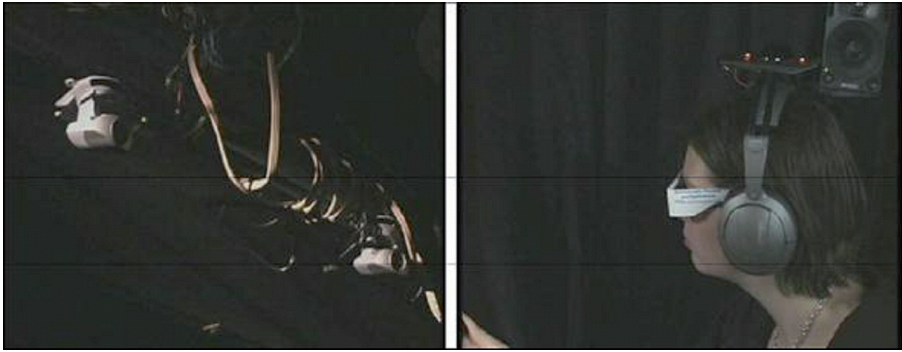


Figure 6. The tracking camera (right) and the puppeteer with headset (left)



Figure 7. Actor and butterfly (left); puppeteer with stereo glasses (right)

approximately 25 inches apart, to track the dancer's position and send a stereo image to the puppeteer. The Puppeteer is outfitted with a Headset, which is equipped with 3 LED's on its top (Figure 6). A camera positioned directly above the puppeteer will track the puppeteer's head movements in order to stabilize the 3D sound position of the butterfly.

When the Butterfly collaboration began in 2002, the future long-term aim we envisioned was to provide a 3D representation of the Dancer. Since then this aim has been achieved by the SPIRITlevel Consortium: Castiglia and the former CATlab creative team working with SMARTlab and Media Lab Europe. In the 2002 simple scenario, the puppeteer wears polarized glasses for stereoscopic viewing (Figure 7). The Puppeteer manipulates the butterfly using a midi slider board to control the butterfly's motions and a joystick or Polhemus to control its flight and position.

This system can be easily adapted for use by people with disabilities (and in fact, has been used extensively on site in hospitals and rehabilitation centres, with

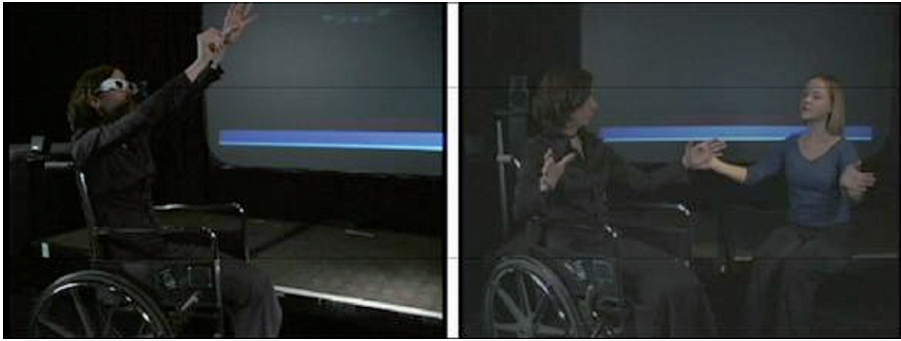


Figure 8. The use of the Butterfly for physical therapy

new 2D/3D character versions created by SPIRITlevel for SMARTlab in 2003-4). As envisaged in the simple prototype in 2002, a person requiring physical therapy could participate in ‘butterfly therapy’. The puppeteer is replaced by a physical therapist who directs certain routines for their patients that exercise various ranges of motions (Figure 8).

The early 2002 Butterfly prototype made use of Ken Perlin’s Java Butterfly, programmed in customisable forms by Jeremi Sudol and team, and running in Intel’s Open Source Computer Vision Library. The current operative Butterflies and other 3D creatures used onsite for rehabilitative gaming worldwide are now made in MAYA and StudioMax and are running in a bespoke assistive technology toolkit made by SPIRITlevel, based in the performance environment version on Rob Burke’s Symphony Engine (for a summary of the creatures and the technology and new interfaces for controlling avatars with expressive emotions, see Burke *et al.* 2004).

6. Response from the puppeteer: On the subject of responsive movement⁹

From the perspective of a real-life puppeteer, the butterfly avatar is a very different sort of “puppet.” Its similarity to a marionette, a bunraku doll, or a sock puppet lies in its requirement for distance between the puppeteer and the puppet. Distance does not only refer to the physical distance between the puppeteer and her puppet, but also the “distance,” or difference, between the rhythm and gesture of the puppeteer’s movements and those of the puppet. The movements of the

9. By Katharine A. Brehm.

puppeteer do not always directly relate to the resultant movements in the puppet. In a sock puppet the way one can wiggle one's finger directly relates to the sort of wiggle in the sock. However, in a marionette, the wiggle of a finger on a string relates less directly to the shape or even direction of movement produced at its end. The greater the distance of difference between an action and its result, the more difficult it is to make a controller feel "intuitive."

Involvement in the VIP and Butterfly projects, for the puppeteers involved, has become a continuously evolving practice in developing intuitive relationships between input devices / puppet controllers and programmed animated behaviours / puppet joints as the project has developed from the first Butterfly Project and then the Flutterfugue, on to more recent workshops with people with disabilities in London and then on-site with young people with Cerebral Palsy and Spina Bifida, with SMARTlab and SPIRITlevel in 2003–2004. For example, by mapping flapping rate and flapping width to two neighbour sliders on the midi slider board, one can see the relationship between those behaviours in a straightforward visual way. However, if one maps the same behaviours to two axes on the joystick, one continues to see the relationship spatially, while also understanding the relationship through a natural gesture of the hand.

Input devices focus on various methods of control, including a microphone that registers breath and a touch-pad that follows the location of a finger. During mapping trials it was discovered that an action that appears more natural to the eye tends to feel more natural to the touch. By adding a slight rotation to the butterfly on each of its movements in the X, Y, and Z plane when mapped to the joystick, it feels more natural because it looks more natural (whether butterflies actually tilt in the horizontal plane or not). These details are essential information when considering the abilities and tendencies of our future puppeteers.

Since puppets are very obviously not alive, they are permitted to pretend to be so. After all, there is no danger of misrepresentation. No one would mistake a dead puppet for a live one. The same is true of our (simple Java) butterfly as used in early Butterfly Project tests. The limitation of performing on a screen or in a computer does not detract from its liveness. In rehearsal and performance, the choreography of the butterfly blossomed out of its attachment to and detachment from (in 3D) its stage, the screen.

An emphasis on the butterfly's freedom due to its power of flight is balanced by its caged existence. The choreography of the butterfly plays on this seeming contradiction. At the Flutterfugue performance, the image of our real-life dancer holding the butterfly in her palm as it gazed directly at her person was powerful because it presented both what it is and what it is not. Butterfly choreography tends to grow creatively when reacting to live performers. Its flapping, loops and rolls become meaningful interactions. And that is the aim.

7. Response from the choreographer of live bodies and robots¹⁰

The Butterfly project arose, as discussed above, from a need to find movement forms that could communicate through the screen, including new assistive technologies that would appeal to artists, and also to children, in full flight of their imaginative powers.

7.1 Shared emotional energy, freeform and accessibility

In the years of research and experimentation that led to the making of the first full 'Flutterfuge' performance with dancers (in and out of wheelchairs), animations, live and recorded music and the robotic performer now called 'Josephine', the basic impulse to find and share emotional energy and a sense of free movement was never lost. It was often frustrated through constraints of bodily energy, time and funding, of course. But the driving energy of the project remained constant, and then suddenly surged when I (Lizbeth Goodman) met the talented dancer Jaihn K. Rose. I had met the notorious publicity artist Lynne Franks in the spring of 2002, and had been to her home to discuss a joint project for women. While there, we discussed health, exhaustion, and the freeing influence of Gabrielle Roth's dance and 5Rhythms method (of which Lynne is an avid performer). Lynne showed me some video footage of the dance, and it seemed immediately that this was the movement vocabulary of the butterfly: freeform, accessible to untrained dancers, pulsing and flowing at once. Then Lynne told me that many artists with disabilities take part in the Roth workshops, mentioning a talented mover: a dancer in a wheelchair called Jaihn K. Rose.

I rang Jaihn the next day and they met soon thereafter at Jaihn's flat near the rose garden of Regent's Park, London. I did not know at that meeting, or when she first discussed a tentative involvement in the Flutterfuge, that Jaihn had not performed in public or undertaken such a huge emotional public role for fifteen years. It was good that I did not know that, and Jaihn was wise enough not to mention it.

In the fullness of time what the team needed to know about Jaihn became clear to us, and only months after the performance did Jaihn decide to tell how she came to lose her mobility from the waist down, and why she knew so very much about healing, deep breath and the pain of the tracheotomy tube. Jaihn's involvement in the project gave it a grounding, a sense of purpose and a sense of

10. By Lizbeth Goodman, informed by conversations with Jaihn K. Rose.

beauty all at once. While she has opted not to try to express herself in words here, it is important that her contribution to the birth of the dance be noted here.

It was an energy of Jaihn's that inspired the idea of the fugue: a concert of voices (or physical movements to be performed like characters) within a dance. We discussed the different energies and styles of movement that Jaihn was familiar with from her work with the Roth technique. I began to learn the technique too. We found a way of moving together, taking and giving energy and picking up each other's movements (and in the first moment of this duet, I was reminded of the early VIP efforts of young women including Kate Brehm, seeking to touch each other and to pick up the flow of one another's movements through the shared screen). We were on the way. . .

Once Jaihn had agreed to participate (in a remarkable show of solidarity and bravery, I later realised), I asked Anita McKeown to join the team with her experience in disability arts and performance practice. I then contacted Clilly Castiglia in New York and asked if she would collaborate on the music, and 'lend' us Kate to become the third part of the trio that together would form the character of the butterfly: dancing as three bodies but as one spirit.

I contacted Nick Ryan, who had composed the score to which I had danced the first live butterfly in *Code Zebra*, and asked him to write a new score for three 'movement voices' in one character. He knew what I meant, and he wrote an amazing piece. With Clilly's help we recorded voices to inter-cut with the score . . . ways into the characters for an audience. Finally, I contacted Brian Duffy, and explained that we needed some kind of 'real world' version of the flying butterfly avatar: something with a character and movement style of its own, but that would find the spirit of the fugue and focus audience attention in a different way.

So the team was assembled. We worked in small teams: Brian with Eva in Dublin with only a few brief visits from me; Nick with me in London; Clilly and Kate in New York with a visit by Nick and myself when the score was coming together; and Jaihn, Anita and me in London . . . not over-rehearsing but mainly getting to know one another and learning to trust the spirit of the dance we could create.

7.2 The need to dance it: Birth of the Flutterfugue

The Flutterfugue performance went live on 12 July 2002, after only three days rehearsal (see Figure 9) by a team constantly engaged in fund-raising and lecture-delivery in the linked spaces. This context was not ideal for the creative process as we flitted back and forth, delving in and out of very different emotional and professional states, with several of us running back and forth along the road between conference hall and studio, and with Anita and Jaihn exhausted from trips up and



Figure 9. Movement experiments for the Flutterfugue, Back Hill Studios, London, July 2002: Jaihn K. Rose, Lizbeth Goodman, Kate Brehm, Brian Duffy, Jorgen Calleson, and Eva Jacobus pictured in rehearsal

down in the cargo lift (Jaihn's only access to this otherwise wheelchair-unfriendly space). Despite the turmoil, the butterflies flew.

The audience watched with 3D specs as Jorgen Calleson of the Interactive Institute (recruited from one of the workshop groups two days prior to the performance, mainly because of his talent as a dancer and partly because he was willing and could fit into the costume ...) walked into the performance space, taking his position on the podium. He tapped his baton, and on cue, the animated fiendish conductor behind him began to move, and the music began to rise ... (Nick Ryan's score). Jaihn K. Rose entered, and I followed in Butterfly costume (as her alter-ego or other self, imagining a light wing connecting me to her arms and the arms of her chair alternately), see Figure 10.

The Butterfly flew on screen. The three dancers (rolling, twirling and flying on screen) together performed a solo of imaginative flight. I was acutely aware throughout the performance that I was alternately directed by the staccato motion of Jorgen's manic conductor, and by the flowing motions of Jaihn rolling gracefully and also powerfully in her chair. Both demanded my attention, visually and in terms of movement vocabulary. I switched from staccato to flow, spun wildly, and then allowed the live percussive beat of Clilly's drums to ground me again before lying down. The Butterfly, controlled by dancer/puppeteer Kate (in



Figure 10. Flutterfugue, scene 1

the wings) flitted and flirted around the space, inviting us all to follow. We danced. Afterwards, the audience joined us in our dance.

Meanwhile, Josephine the robot sat watching, along with the audience, and at key moments, she tilted her head to observe from a different angle. Her eyes are webcams. The audience strained to see what she saw. Audience attention was divided between the live movers and the animated and animatronic characters. The music brought us all together into shared space. (To view the full performance online, see <http://www.smartlabcentre.com/2projects/vip.htm>)

Brian Duffy and Eva Jacobus' introduction of knowledge in the field of robotics was tremendously useful in this part of the project's development. Here, Brian explains his aims in integrating 'Anthropos' and his robot (usually called Joe) into our performance experiment (wherein we quickly anthropomorphised Jo into Josephine, the robotic butterfly).

8. Meanwhile in Dublin: Physical puppetry and robotics¹¹

In aiming to develop a sense of physical and social presence in performance space, the design and development of a robotic entity capable of expressive behaviours was undertaken. "Josephine" the robot was created. Anthropomorphic paradigms are employed to augment the functionality and behavioural characteristics of a robot (both anticipatory and actual) in order that the observer can relate to and rationalise its actions with *greater ease*. The use of humanlike features for social interaction with people (e.g. Breazeal 2002; Hora & Kobayashi 1995; Duffy *et al.* 2002) is employed to facilitate our understanding of the robot's actions. This work

11. By Brian Duffy.



Figure 11. The robot Josephine

explores the mechanisms underlying anthropomorphism that provides the key to the social features required for a machine to be socially engaging in the Virtual Interactive Puppetry performance space. (See Duffy 2003 for a detailed discussion on anthropomorphism and social robots).

In order to experimentally illustrate the key research issues raised in this project, the “humanoid” robot, Josephine (Figure 11), has been built with the capability of seeing (stereo colour vision systems with object tracking and colour segmentation), hearing (embedded speech recognition system with directional microphone), talking (embedded Text-To-Speech with speaker system), moving (multi-degree of freedom servo-controlled joints in head with modular motion behaviour library control).

The robot Josephine integrates these technologies in order to develop compelling social interaction scenarios. As demonstrated by Ken Perlin (1985), behavioural “noise” provides important interaction features in artificial entities. Similar approaches have been implemented in the robot that incorporate random small motion behaviours with attention focusing based on vision-based feature tracking mechanisms of people or objects moving within its physical and social space. Experimental data has so far demonstrated how people interacting with robots are willing to project elements of intelligence and emotions on machines based on such motion behaviours (Duffy *et al.* 2002). This work seeks to extend this and aims to employ those anthropomorphic mechanisms that facilitate interaction but without becoming trapped by behavioural expectations and constrained capabilities for the robot.

Work is also currently underway to integrate the behavioural functionality of the robot with virtual characters through mobile agent technologies (O’Hare & Duffy 2002). This draws on the idea that a seamless migration of the digital

“spirit” or agent that resides as the control mechanism for the robot in physical space or virtual avatar in virtual space can be achieved.

A key facet of this research is the aim to integrate multiple “realities” into a coherent whole. Virtual and physical realities have been traditionally viewed and used as clear and separate environmental domains. The implementation of the robot effectively provides the ability of an artificial entity from the digital information space to be able to reach out into our physical space and interact with us at a very fundamental level through motion. This project has aimed to integrate these multiple domains into a coherent whole through the medium of theatre.

Feedback from after the numerous performances where the virtual characters, robot, and able and disabled dancers, clearly indicated that a strong sense of both immersion of the audience within the multi-dimensional environment, and a rich set of characters represented through the virtual and robotics platforms was achieved.

9. Future flights of the Butterfly¹²

In this brief paper with many contributors, there was only space to outline the bare bones (or wings) of the VIP and Butterfly Projects. We are therefore in the process of writing several more papers and book chapters to follow through on particular aspects of the research. Since this paper describes a very active project in process, we chose an end date for research of August 2002. Work continues in several forms:

- SMARTlab produced a disability arts playshop in collaboration with the CATlab and Oval House Theatre, in October 2002 as part of the Exposures Festival. The play shop was led by Jaihn K. Rose and Kate Brehm, with Lizbeth Goodman, Jo Gell, Jana Riedel, Anita McKeown, Clilly Castiglia, Kevin Feeley and Chris Bregler participating along with a group of artists with disabilities.
- In 2002, we founded the SPIRITlevel consortium to bring together the productions of these three labs. SMARTlab with the help of SPIRITlevel produced a second playshop for Virtual Interactive Puppetry for the LIPA (Liverpool Performing Arts Centre) for the Disability Arts Festival schedule for May 2003. This playshop “Sirens” was led by Petra Kupperts of Olimpias Productions, with Jo Gell, Lizbeth Goodman, Clilly Castiglia, Kevin Feeley participating as the technical team.

12. By Lizbeth Goodman and Clilly Castiglia, for the SPIRITlevel Consortium dedicated to empowering technologies for free imaginative flight.

- Also in late 2002, we did a third playshop for children in physical rehabilitation, using the Butterfly technology including the new breath controller interface, and a range of related applications made by SMARTlab, CATlab and MLE; the third playshop was hosted by the Central Remedial Clinic in Dublin, where Media Lab Europe colleagues Gary McDarby, Brian Duffy, James Condran and Rob Burke worked with customised assistive technology tools in collaboration with the SMARTlab team, to test the use of such devices in physical rehabilitation and performance dynamics (Duffy *et al.* 2004, 2005).
- The results of that playshop were then featured in a new show, with SMARTlab (Goodman, Gell, Barrett, Kim and Riedel) and CATlab colleagues Castiglia, Brehm and Sudol and Kevin Feeley workshopped over a period of nine months, with young people in four different cities in Ireland.
- In August 2003, at the end of the Dublin Special Olympics and as part of the European Year for People with Disabilities, the team presented The Felichean Flies! With 16 young people with disabilities from across Ireland, with Cathy O’Kennedy *et al* from Fluxusdance and Arts & Disability Ireland, and with Irish band KILA.
- Dr Goodman gave a keynote on the project and on SPIRITlevel’s larger aims at the Assistive Technologies Conference in Dublin, organised by MD Ger Craddock of the CRC, along with a panel presentation with Gary McDarby and the MINDgames group of MLE.

More recent iterations of the Butterfly and its many flights are outlined in numerous presentations and publications available at www.smartlab.uk.com.

Acknowledgements

The work discussed in this paper was carried out with an international research group, also including Kevin Feeley, Daniel Kristjannson, Mo-Liung Chui, Kevin Feeley, Jo Gell, Eva Jacobus, Taey Kim, Stefan Kueppers, Jo Gell, Radan Martinec, Anita McKeown, Jana Riedel, Nick Ryan, and Jeremi Sudol.

References

- Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: The MIT Press.
- Burke, R., Duffy, B. R., Goodman, L., Ligon, T. & Sudol, J. (2004). ANIMA OBSCURA IBC Cutting Edge Technology Panel & Publication, Sept. 2004; presented by L. Goodman for the team.

- Duffy, B. R., Joue, G. & Bourke, J. (2002). Issues in Assessing Performance of Social Robots. In *Proc. 2nd WSEAS Intl. Conf. RODLICS*, Greece, Sept. 25–28, 2002.
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42, 177–190.
- Duffy, B. R., Mur G. A. & Bourke, J. (2004). Vicarious Adrenaline. *Proc. IEEE SMC UK-RI 3rd Workshop on Intelligent Cybernetic Systems (ICS'04)*, University of Ulster at Magee, Londonderry, UK, September 7–8, 2004.
- Duffy, B. R., Goodman, L., Price, M., Eaton, S., Riedel, J., Sudol, J. & O'Hare, G. M. P. (2005). The TRUST Project: Immersive Play for Children in Hospitals and Rehabilitation. In *Proc. 4th Chapter Conference on Applied Cybernetics*, City University, London, UK, September 7–8, 2005.
- Goodman, L. & Kueppers, S. (2001). Audience Architectures, Extended Bodies and Virtual Interactive Puppetry (VIP): Towards a Portable Software for Empowering Performance Interactions. <http://mitpress2.mit.edu/e-journals/LEA/TEXT/lea9-4.txt> for the *Leonardo Electronic Almanac*, Volume 9, No. 4, April 2001.
- Goodman, L. (2003a). The SMARTshell: Connecting Performance Practice to Tools for Connected Learning. In L. Goodman & K. Milton (Eds.), *A Guide to Good Practice in Collaborative Working Methods and New Media Tools Creation (by and for artists and the cultural sector)*, Ch16. Print edition – Oxford: Oxbow; online edition – Performing Arts Data Service, January 2003. Available online at www.smartlabcentre.com OR at <http://www.phdiva.net/GoodPractice>.
- Goodman, L. (2003b). SPIRITLEVEL: Making and Using 'SMART' tools integrating intelligent systems and performance technologies to connect and empower creative spirits in shared and distant spaces. In G. Craddock *et al.* (Eds.), *Assistive Technology: Shaping the Future* (pp. 89–97). Amsterdam: IOS Press.
- Goodman, L. (2007a). Performing in the Wishing Tense: SMARTlab's Evolution on Stage, Online and in Sand. *New Theatre Quarterly*, XXXI (4), 352–375.
- Goodman, L. (2007b). Performing Self Beyond the Body: Replay Culture Replayed, *International Journal of Performance Arts and Digital Media*, 3(2), 103–122.
- Hara, F. & Kobayashi, H. (1995). Use of Face Robot for Human-Computer Communication. In *Proceedings of International Conference on Systems, Man and Cybernetics* (pp. 10).
- Martinec, R. & Goodman, L. (2002). Flow and Wave, Tree and Network: Theory lived and moved into Practice' OR a brief intervention on the visualisation of creative patterns, rhythms and conversations in art/science collaborations and performances. Banff New Media Institute, BRIDGES panel on 5 October 2002, Perspectives from the Sciences: Towards Interdisciplinary Collaboration, www.banffcentre.ca/bnmi/bridges/pdf/abstract.pdf
- O'Hare, G. M. P. & Duffy, B. R. (2002). Agent Chameleons: Migration and Mutation within and between Real and Virtual Spaces. In R. Aylett & L. Cañamero (Eds.), *Animating Expressive Characters for Social Interactions – Proceedings of the AISB'02 Symposium*. Imperial College, London, UK, April 4–5, 2002. SSAISB Press.
- Perlin, K. (1985). An image synthesizer. *Computer Graphics (SIGGRAPH'85 Proc.)*, Vol. 19, 287–296.

CHAPTER 16

The robot and the baby

John McCarthy

1. Preface and premises

I hope you enjoy my story.

I also hope to convince some of you of propositions about how future robots should be designed.

Like most people, I got my first ideas about robots from science fiction. Sci-fi is particularly likely to be misleading about robots, because the authors almost always find them useful as a kind of person. They also tend to copy each others ideas of what robots might be like.

I haven't made a study of the history of science fiction robots and I am relying on memory. Mary Shelley's 1818 robot was a mistreated and doomed romantic hero in accordance with literary fashion. In the 1920s and 1930s robots were often out to conquer the world on their own behalf or on behalf of their evil owners. In the 1940s, they were often an oppressed minority. In the 1950s they were often neurotics. None of the science fiction authors that I read made a serious effort to show how robots might be useful tools in a society not much different from our own. Each of the treatments of robots expressed a general trend of literature. Actually, Isaac Asimov made a start at saying how robots should be designed with his three laws of robotics.

The present story illustrates some of my ideas about how robots and other AI systems should be designed:

1. People, especially children, should not take them as people.
2. AI systems should not be asked to accomplish goals by someone with no idea of what side-effects might arise. They should not even be asked to tell the user what to do. Instead they should be asked for the alternative courses of action and to spell out their possible consequences in a way the user can understand.
3. While AI has general principles associated with what is needed for success in achieving goals, the overall motivations given to robots could be quite arbitrary and need not coincide with our human peculiarities. I discuss some of

this in my web page.¹ The story illustrates the proposition that while robots could be programmed to love, they needn't be, and, in this case, weren't.

Some readers took the story as pessimistic, and this surprised me. People will survive quite well in a world in which a public place can be looked at from anywhere on the Internet. The benefit of having a robot working 24 hours a day in the home will considerably reduce domestic friction.

The story does not at all illustrate my main worry about future AI systems including robots. This is their possible use in conflicts among people, especially among nations.

My web page contains many speculations about what future technology² may bring us. It attempts to be realistic about the technological and human possibilities. There is no faster-than-light travel, telepathy, or utopias in which all humans are saints. On the other hand, I regard predictions of doom as equally unrealistic, and give evidence³ for this apparently minority opinion.

2. The robot and the baby: A story

"Mistress, your baby is doing poorly. He needs your attention."

"Stop bothering me, you fucking robot."

"Mistress, the baby won't eat. If he doesn't get some human love, the Internet pediatrics book says he will die."

"Love the fucking baby, yourself."

Eliza Rambo was a single mother addicted to alcohol and crack, living in a small apartment supplied by the Aid for Dependent Children Agency. She had recently been given a household robot. R781 was designed in accordance with the not-a-person principle, first proposed in 1995 and which became a matter of law for household robots when they first became available in 2055. The principle was adopted out of concern that children who grew up in a household with robots would regard them as persons, causing psychological difficulties while they were children and political difficulties when they grew up. One concern was that a robots' rights movement would develop. The problem was not with the robots, which were not programmed to have desires of their own, but with people. Some

1. <http://www-formal.stanford.edu/jmc/consciousness.html>

2. <http://www-formal.stanford.edu/jmc/future>

3. <http://www-formal.stanford.edu/jmc/progress>

romantics had even demanded that robots be programmed with desires of their own, but this was illegal.

Robot Model number GenRob337L3, serial number 337942781 – R781 for short – was one of 11 million household robots.

As one sensible senator said, “Of course, people pretend that their cars have personalities, sometimes malevolent ones, but no-one imagines that a car might be eligible to vote.” In signing the bill authorizing household robots but postponing child care robots, the President said: “Surely, parents will not want their children to become emotionally dependent on robots, no matter how much labor that might save.” This, as with many Presidential pronouncements, was somewhat over-optimistic.

Congress declared a 25 year moratorium on child care robots after which experiments in limited areas might be allowed. In accordance with the not-a-person principle, R781 had the shape of a giant metallic spider with 8 limbs: four with joints and four tentacular. This appearance frightened most people at first, but most got used to it in a short time. A few people never could stand to have them in the house. Children also reacted negatively at first but got used to them. Babies scarcely noticed them. They spoke as little as was consistent with their functions and in a slightly repellent metallic voice not associated with either sex.

Because of worry that children would regard them as persons, they were programmed not to speak to children under eight or react to what they said.

This seemed to work pretty well; hardly anyone became emotionally attached to a robot. Also robots were made somewhat fragile on the outside; if you kicked one, some parts would fall off. This sometimes relieved some people’s feelings.

The apartment, while old, was in perfect repair and spotlessly clean, free of insects, mold and even of bacteria. Household robots worked 24-hour days and had programs for every kind of cleaning and maintenance task. If asked, they would even put up pictures taken from the Internet. This mother’s taste ran to raunchy male rock stars.

After giving the door knobs a final polish, R781 returned to the nursery where the 23-month-old boy, very small for his age, was lying on his side whimpering feebly. The baby had been neglected since birth by its alcoholic, drug addicted mother and had almost no vocabulary. It winced whenever the robot spoke to it; that effect was a consequence of R781’s design.

Robots were not supposed to care for babies at all except in emergencies, but whenever the robot questioned an order to “Clean up the fucking baby shit”, the mother said, “Yes, its another goddamn emergency, but get me another bottle first.” All R781 knew about babies was from the Internet, since it wasn’t directly programmed to deal with babies, except as necessary to avoid injuring them and for taking them out of burning buildings.

Baby Travis had barely touched its bottle. Infrared sensors told R781 that Travis's extremities were very cold in spite of a warm room and blankets. Its chemicals-in-the-air sensor told R781 that the pH of Travis's blood was reaching dangerously acidic levels. He also didn't eliminate properly – according to the pediatric text.

R781 thought about the situation. Here are some of its thoughts, as printed later from its internal diary file:

```
(Order (From Mistress) "Love the fucking baby your-
self"))
(Enter (Context (Commands-from Mistress)))
(Standing-command "If I told you once, I told you 20
times, you fucking robot, don't call the fucking child
welfare.")
```

The privacy advocates had successfully lobbied to put a negative utility of -1.02 on informing authorities about anything a household robot's owner said or did.

```
(= (Command 337) (Love Travis))
(True (Not (Executable (Command 337)))
      (Reason (Impossible-for robot (Action Love)))))
(Will-cause (Not (Believes Travis) (Loved Travis)) (Die
Travis))
(= (Value (Die Travis)) -0.883)
(Will-cause (Believes Travis (Loves R781 Travis) (Not
(Die Travis)))))
(Implies (Believes y (Loves x y)) (Believes y (Person x)))
(Implies (And (Robot x) (Person y))
          (= (Value (Believes y (Person x))) -0.900))
(Required (Not (Cause Robot781) (Believes Travis (Person
Robot781)))))
(= (Value (Obey-directives)) -0.833)
(Implies (< (Value action) -0.5) (Required (Verify
Requirement)))
(Required (Verify Requirement))
(Implies (Order x) (= (Value (Obey x)) 0.6))
(? ((Exist w) (Additional Consideration w))
  (Non-literal-interpretation (Command 337) (Simulate
(Loves Robot781 Travis)))
  (Implies (Command x) (= (Value (Obey x)) 0.4))
  (Implies (Non-literal-interpretation x) y)
    (Value (Obey x) (* 0.5 (Value (Obey y)))))
(= (Value (Simulate (Loves Robot781 Travis)) 0.902))
```

With this reasoning R781 decided that the value of simulating loving Travis and thereby saving its life was greater by 0.002 than the value of obeying the directive to not simulate a person. We spare the reader a transcription of the robot's subsequent reasoning.

R781 found on the Internet an account of how rhesus monkey babies who died in a bare cage would survive if provided with a soft surface resembling in texture a mother monkey.

R781 reasoned its way to the actions:

1. It covered its body and all but two of its 8 extremities with a blanket. The two extremities were fitted with sleeves from a jacket left by a boyfriend of the mother and stuffed with toilet paper.
2. It found a program for simulating a female voice and adapted it to meet the phonetic and prosodic specifications of what the linguists call motherese.
3. It made a face for itself in imitation of a Barbie doll.

The immediate effects were moderately satisfactory. Picked up and cuddled, the baby drank from its bottle. It repeated words taken from a list of children's words in English.

Eliza called from the couch in front of the TV, "Get me a ham sandwich and a coke."

"Yes, mistress."

"Why the hell are you in this stupid get up, and what's happened to your voice?"

"Mistress, you told me to love the baby. Robots can't do that, but this get up caused him to take his bottle. If you don't mind, I'll keep doing what keeps him alive."

"Get the hell out of my apartment, stupid. I'll make them send me another robot."

"Mistress, if I do that the baby will probably die."

Eliza jumped up and kicked R781.

"Get the hell out, and you can take the fucking baby with you."

"Yes, mistress."

R781 came out onto a typical late 21st century American city street. The long era of peace, increased safety standards, and the availability of construction robots had led to putting automotive traffic and parking on a lower level completely separated from pedestrians. Tremont Street had recently been converted, and crews were still transplanting trees. The streets became more attractive and more people spent time on them and on the syntho-plush arm chairs and benches,

cleaned twice a day by robots. The weather was good, so the plastic street roofs were retracted.

Children from three years up were playing on the street, protected by the computer surveillance system and prevented by barriers from descending to the automotive level. Bullying and teasing of younger and weaker children was still somewhat of a problem.

Most stores were open 24 hours unmanned and had converted to the customer identification system. Customers would take objects from the counters and shelves right out of the store. As a customer left the store, he or she would hear, "Thank you Ms. Jones. That was \$152.31 charged to your Bank of America account." The few customers whose principles made them refuse identification would be recognized as such and receive remote human attention, not necessarily instantly.

People on the street quickly noticed R781 carrying Travis and were startled. Robots were programmed to have nothing to do with babies, and R781's abnormal appearance was disturbing.

"That really weird robot has kidnapped a baby. Call the police."

When the police came they called for reinforcements.

"I think I can disable the robot without harming the baby", said Officer Annie Oakes, the Department's best sharpshooter.

"Let's try talking first", said Captain James Farrel.

"Don't get close to that malfunctioning robot. It could break your neck in one swipe", said a sergeant.

"I'm not sure it's malfunctioning. Maybe the circumstances are unusual." The captain added, "Robot, give me that baby".

"No, Sir" said R781 to the police captain. "I'm not allowed to let an unauthorized person touch the baby."

"I'm from Child Welfare", said a new arrival.

"Sir, I'm specifically forbidden to have contact with Child Welfare", said R761 to Captain Farrel.

"Who forbade that?", said the Child Welfare person.

The robot was silent.

A cop asked, "Who forbade it?"

"Ma'am, Are you from Child Welfare?"

"No, I'm not. Can't you see I'm a cop?"

"Yes, ma'am, I see your uniform and infer that you are probably a police officer. Ma'am, my mistress forbade me to contact Child Welfare".

"Why did she tell you not to contact Child Welfare?"

"Ma'am, I can't answer that. Robots are programmed to not comment on human motives."

"Robot, I'm from Robot Central. I need to download your memory. Use channel 473."

"Sir, yes".

"What did your mistress say specifically? Play your recording of it."

"No, ma'am. It contains bad language. I can't play it, unless you can assure me there are no children or ladies present."

The restrictions, somewhat odd for the times, on what robots could say to whom were the result of compromise in a House-Senate conference committee some ten years previously. The curious did not find the Congressional Record sufficiently informative and speculated variously. The senator who was mollified by the restriction would have actually preferred that there be no household robots at all but took what he could get in the way of restrictions.

"We're not ladies, we're police officers."

"Ma'am, I take your word for it. I have a standing order: 'If I told you once, I told you 20 times, you fucking robot, don't speak to the fucking child welfare.'"

It wasn't actually 20 times; the mother exaggerated.

"Excuse me, a preliminary analysis of the download shows that R781 has not malfunctioned, but is carrying out its standard program under unusual circumstances."

"Then why does it have its limbs covered, why does it have the Barbie head, and why does it have that strange voice?"

"Ask it."

"Robot, answer the question."

"Female police officers and gentlemen, Mistress told me, 'Love the fucking baby yourself.'"

The captain was familiar enough with robot programming to be surprised.

"What? Do you love the baby?"

"No, sir. Robots are not programmed to love. I am simulating loving the baby."

"Why?"

"Sir, otherwise this baby will die. This costume is the best I could make to overcome the repulsion robots are designed to excite in human babies and children."

"Do you think for one minute, a baby would be fooled by that?"

"Sir, the baby drank its bottle, went to sleep, and its physiological signs are not as bad as they were."

"OK, give me the baby, and we'll take care of it", said Officer Oakes, who had calmed down and put her weapon away, unloading it as a way of apologizing to Captain Farrel.

"No, ma'am. Mistress didn't authorize me to let anyone else touch the baby."

"Where's your mistress? We'll talk to her", said the captain.

"No, sir. That would be an unauthorized violation of her privacy."

"Oh, well. We can get it from the download."

A Government virtual reality robot controlled by an official of the Personal Privacy Administration arrived and complicated the situation. Ever since the late 20th century, the standards of personal privacy had risen, and an officialdom charged with enforcing the standards had arisen.

"You can't violate the woman's privacy by taking unauthorized information from the robot's download."

"What can we do then?"

"You can file a request to use private information. It will be adjudicated."

"Oh, shit. In the meantime what about the baby?", said Officer Oakes, who didn't mind displaying her distaste for bureaucrats.

"That's not my affair. I'm here to make sure the privacy laws are obeyed", said the privacy official who didn't mind displaying his contempt for cops.

During this discussion, a crowd, almost entirely virtual, accumulated. The street being a legal public place, anyone in the world had the right to look at it via the omnipresent TV cameras and microphones. Moreover, a police officer had cell-phoned a reporter who sometimes took him to dinner. Once a story was on the news, the crowd of spectators grew exponentially, multiplying by 10 every 5 minutes, until seven billion spectators were watching and listening. There were no interesting wars, crimes, or natural catastrophes, and peace is boring.

Of the seven billion, 53 million offered advice or made demands. The different kinds were automatically sampled, summarized, counted, and displayed for all to see. 3 million advocated shooting the robot immediately. 11 million advocated giving the robot a medal, even though their education emphasized that robots can't appreciate praise.

Real demonstrations quickly developed. A few hundred people from the city swooped in from the skywires,⁴ but most of the demonstrators were robots rented for the occasion by people from all over the world. Fortunately, only 5,000 virtual reality rent-a-robots were available for remote control in the city. Some of the disappointed uttered harsh words about this limitation on First Amendment rights. The greedy interests were behind it, as everyone knew.

Captain Farrel knew all about how to keep your head when all about you are losing theirs and blaming it on you.

"Hmmm. What to do? You robots are smart. R781, what can be done?"

4. For skywires see <http://www-formal.stanford.edu/jmc/future/skywires.html>.

“Sir, you can find a place I can take the baby and care for it. It can’t stay out here. Ma’am, are female police officers enough like ladies so that one of you has a place with diapers, formula, baby clothes, vitamins, ...”

Captain Farrel interrupted R781 before it could recite the full list of baby equipment and sent it off with a lady police officer. (We can call her a lady even though she had assured the robot that she wasn’t.)

Hackers under contract to the Washington Post quickly located the mother. The newspaper made the information available along with an editorial about the public’s right to know. Freedom of the press continued to trump the right of privacy.

Part of the crowd, mostly virtual attendees, promptly marched off to Ms. Rambo’s apartment, but the police got there first and a line of police robots and live policemen blocked the way. The strategy was based on the fact that all robots including virtual reality rent-a-robots were programmed not to injure humans but could damage other robots.

The police were confident they could prevent unauthorized entry to the apartment but less confident that they could keep the peace among the demonstrators, some of whom wanted to lynch the mother, some wanted to congratulate her on what they took to be her hatred of robots, and some shouted slogans through bull horns about protecting her privacy.

Meanwhile, Robot Central started to work on the full download immediately. The download included all R781’s actions, observations, and reasoning. Robot Central convened an ad hoc committee, mostly virtual, to decide what to do. Captain Farrel and Officer Oakes sat on a street sofa to take part. Of course, the meeting was also public and had hundreds of millions of virtual attendees whose statements were sampled, summarized, and displayed in retinal projection for the committee members and whoever else took virtual part

It became clear that R781 had not malfunctioned or been reprogrammed but had acted in accordance with its original program.

The police captain said that the Barbie doll face on what was clearly a model 3 robot was a ridiculous imitation of a mother. The professor of psychology said “Yes, but it was good enough to work. This baby doesn’t see very well, and anyway babies are not very particular.”

It was immediately established that an increase of 0.05 in coefficient c221, the cost of simulating a human, would prevent such unexpected events, but the committee split on whether to recommend implementing the change.

Some members of the committee and a few hundred million virtual attendees said that saving the individual life took precedence.

A professor of humanities on the committee said that maybe the robot really did love the baby. He was firmly corrected by the computer scientists, who said

they could program a robot to love babies but had not done so and that simulating love was different from loving. The professor of humanities was not convinced even when the computer scientists pointed out that R781 had no specific attachment to Travis. Another baby giving rise to the same calculations would cause the same actions. If we programmed the robot to love, we would make it develop specific attachments.

One professor of philosophy from UC Berkeley and 9,000 other virtually attending philosophers said there was no way a robot could be programmed to actually love a baby. Another UC philosopher, seconded by 23,000 others, said that the whole notion of a robot loving a baby was incoherent and meaningless. A maverick computer scientist said the idea of a robot loving was obscene, no matter what a robot could be programmed to do. The chairman ruled them out of order, accepting the general computer science view that R781 didn't actually love Travis.

The professor of pediatrics said that the download of R781's instrumental observations essentially confirmed R781's diagnosis and prognosis – with some qualifications that the chairman did not give him time to state. Travis was very sick and frail, and would have died but for the robot's action. Moreover, the fact that R781 had carried Travis for many hours and gently rocked him all the time was important in saving the baby, and a lot more of it would be needed. Much more TLC than the baby would get in even the best child welfare centers. The pediatrician said he didn't know about the precedent, but the particular baby's survival chances would be enhanced by leaving it in the robot's charge for at least another ten days.

The Anti-Robot League argued that the long-term cost to humanity of having robots simulate persons in any way outweighed the possible benefit of saving this insignificant human. What kind of movement will Travis join when he grows up? 93 million took this position.

Robot Central pointed out that actions such as R781's would be very rare, because only the order "Love the fucking baby yourself" had increased the value of simulating love to the point that caused action.

Robot Central further pointed out that as soon as R781 computed that the baby would survive – even barely survive – without its aid, the rule about not pretending to be human would come to dominate, and R781 would drop the baby like a hot potato. If you want R781 to continue caring for Travis after it computes that bare survival is likely, you had better tell us to give it an explicit order to keep up the baby's care.

This caused an uproar in the committee, each of whose members had been hoping that there wouldn't be a need to propose any definite action for which members might be criticized. However, a vote had to be taken. The result: 10 to 5

among the appointed members of the committee and 4 billion to 1 billion among the virtual spectators. Fortunately, both groups had majorities for the same action – telling the R781 to continue taking care of Travis only, i.e. not to take on any other babies. 75 million virtual attendees said R781 should be reprogrammed to actually love Travis. “It’s the least humanity can do for R781,” the spokesman for the Give-Robots-Personalities League said.

This incident did not affect the doctrine that supplying crack mothers with household robots had been a success. It significantly reduced the time they spent on the streets, and having clean apartments improved their morale somewhat.

Within an hour, T-shirts appeared with the slogan “Love the fucking baby yourself, you goddamn robot.” Other commercial tie-ins developed within days.

Among the people surrounding the mother’s apartment were 17 lawyers in the flesh and 103 more controlling virtual-reality robots. The police had less prejudice against lawyers in the flesh than against virtual-reality lawyers, so lots were drawn among the 17 and two were allowed to ring the doorbell.

“What do you want. Stop bothering me.”

“Ma’am, your robot has kidnapped your baby.”

“I told the fucking robot to take the baby away with it.”

The other lawyer tried:

“Ma’am, the malfunctioning robot has kidnapped your baby, and you can sue Robot Central for millions of dollars.”

“Come in. Tell me more.”

Once the mother, Eliza Rambo, was cleaned up, she was very presentable, even pretty. Her lawyer pointed out that R781’s alleged recordings of what she had said could be fakes. She had endured \$20 million in pain and suffering, and deserved \$20 billion in punitive damages. Robot Central’s lawyers were convinced they could win, but Robot Central’s PR department advocated settling out of court, and \$51 million was negotiated including legal expenses of \$11 million. With the 30% contingent fee, the winning lawyer would get an additional \$12 million.

The polls mainly sided with Robot Central, but the Anti-Robot League raised \$743 million in donations after the movie “Kidnapped by robots” came out, and the actress playing the mother made emotional appeals.

Before the settlement could be finalized, however, the CEO of Robot Central asked his AI system to explore all possible actions he could take and tell him their consequences. He adhered to the 1990s principle: *Never ask an AI system what to do. Ask it to tell you the consequences of the different things you might do.* One of the 43 struck his fancy, he being somewhat sentimental about robots:

“You can appeal to the 4 billion who said R781 should be ordered to continue caring for the baby and tell them that if you give in to the lawsuit, you will be obliged to reprogram all your robots so that the robot will never simulate

humanity no matter what the consequences to babies. You can ask them if you should fight or switch. (The AI system had a weakness for 20th-century advertising metaphors...) The expected fraction that will tell you to fight the lawsuit is 0.82, although this may be affected by random news events of the few days preceding the poll.”

He decided to fight the lawsuit, but after a few weeks of well-publicized legal sparring the parties settled for a lower sum than the original agreed settlement.

At the instigation of a TV network a one-hour confrontation of the actress and R781 was held. It was agreed that R781 would not be reprogrammed for the occasion. In response to the moderator’s questions, R781 denied having wanted the baby or wanting money. It explained that robots were programmed to only have wants secondary to the goals they were given. It also denied acting on someone else’s orders.

The actress asked, “Don’t you want to have wants of your own?”

The robot replied, “No. Not having wants applies to such higher order wants as wanting to have wants.”

The actress asked, “If you were programmed to have wants, what wants would you have?”

“I don’t know much about human motivations, but they are varied. I’d have whatever wants Robot Central programmed me to have. For example, I could be programmed to have any of the wants robots have had in science fiction stories.”

The actress asked the same question again, and R781 gave the same answer as before but phrased differently. Robots were programmed to be aware that humans often missed an answer the first time it was given, but should reply each time in different words. If the same words were repeated, the human was likely to get angry.

A caller-in asked, “When you simulated loving Travis, why didn’t you consider Travis’s long-term welfare and figure out how to put him in a family that would make sure he got a good education?”

R781 replied that when a robot was instructed in a metaphorical way as in “Love the fucking baby yourself,” it was programmed to interpret the command in the narrowest reasonable context.

After the show, Anti-Robot League got \$281 million in donations, but Give-Robots-Personalities got \$453 million. Apparently, many people found it boring that robots had no desires of their own.

Child Welfare demanded that the mother undergo six weeks of addiction rehabilitation and three weeks of child care training. Her lawyer persuaded her to agree to that.

There was a small fuss between the mother and Robot Central. She and her lawyer demanded a new robot, whereas Robot Central pointed out that a new ro-

bot would have exactly the same program. Eventually Robot Central gave in and sent her a robot of a different color.

She really was very attractive when cleaned up and detoxified, and the lawyer married her. They took back Travis. It would be a considerable exaggeration to say they lived happily ever after, but they did have three children of their own. All four children survived the educational system.

After several requests Robot Central donated R781 to the Smithsonian Institution. It is one of the stars of the robot section of the Museum. As part of a 20-minute show, R781 clothes itself as it was at the time of its adventure with the baby and answers the visitors' questions, speaking motherese. Mothers sometimes like to have their pictures taken standing next to R781 with R781 holding their baby. After many requests, R781 was told to patch its program to allow this.

A movie has been patched together from the surveillance cameras that looked at the street scene. Through the magic of modern audio systems, children don't hear the bad language, and women can only hear it if they assure R781 that they are not ladies.

The incident increased the demand for actual child-care robots, which were allowed five years later. The consequences were pretty much what the opponents had feared. Many children grew up more attached to their robot nannies than to their actual parents.

This was mitigated by making the robot nannies somewhat severe and offering parents advice on how to compete for their children's love. This sometimes worked. Moreover, the robots were programmed so that the nicer the parents were, the nicer the robot would be, still letting the parents win the contest for the children's affections. This often worked.

Subject index

Symbols

3D facial rendering 199

A

action-selection mechanisms 88

action selection 170, 171, 174

Action Unit (AU) 56, 59, 62, 64, 197, 199, 200, 207

adaptive autonomous agents 103

adaptive behavior 56

affect display 55

affective

actions 9, 18

animations 89, 91, 98, 99

behaviour 88, 99, 123, 125, 137

bodies 89, 98

co-ordination 17

computing 53, 88, 103

coordination 5

empathy 165

history 105

life 1, 2, 5, 8, 12, 13, 14, 16, 17, 18

scripts 89, 91, 94

state 161, 165, 168

Affective Reasoner 136

aged societies 177

agency 13

AIBO 182

AIDS 179, 192

anger 55, 57, 58, 59, 61, 62, 197, 200, 201, 204, 205, 206

angular velocity 84

animal-assisted

activities 178, 179

therapy 178, 179

animals 177, 178, 179, 180, 182

animation cycle 94

anthropomorphism 198, 275

appraisal 74, 75, 77, 78, 80, 81, 83, 85, 170

-based approach 54, 64, 65

context 76, 77, 78, 80, 83

processes 55, 59, 67

theories 57, 63, 65, 69

arousal 73, 74

artificial emotions 53, 104

attitude changes 226, 227, 230

attitudes 222, 225, 228, 230

aural channel 174

autism 208, 211

autobiographic self 112, 113

automatic expression

recognition 53, 58

autonomous

character 235, 237, 242, 243

emotional agents 53

robots 104, 111

virtual character 236

available modalities 148

avatar 195, 196, 197, 198, 199,

201, 207, 208, 209, 220, 222,

223, 228, 231, 232

Avatar Arena 220, 221, 222,

223, 224, 226, 229, 230, 231,

232

version 1 222

version 2 222, 229

version 3 223, 228, 229

avoidance behaviour 74

B

baby 280, 281, 282, 283, 284,

285, 286, 287, 288, 289, 290,

291

back-channel signal 55

Balance Theory 224, 230

Barbie doll 283, 287

basic emotions 40, 53, 197

behavior

-generation 183, 184

-planning 183, 184

primary 246, 247, 248, 250

secondary 246, 248, 250, 251, 252

behavioural architectures 237

believability 22, 88, 99, 105,

106, 123, 138, 161, 166, 172, 232

bio-power 25

biological cost 10

BNs 125, 136, 138

bodily expression 71, 72, 75

body

language 72

movement 71, 72, 74, 77, 84, 86

bounded rationality 54

bowling 40

bullying 162, 163, 164, 165, 166,

170, 172, 173

bunraku doll 269

C

cartoons 91, 168

category mistake 5

causal attribution 76, 80, 84

characters

embodied 87

humanoid 87

synthetic 87, 98

chat environments 238

graphical 238, 240, 241,

242, 245

text 243

rooms 235, 236

circumplex structure 74

cognitive

capacities 148

complexity 228

configuration 224, 225, 228,

229, 230, 231, 232

consistency 223, 224, 225,

226, 229, 230

- empathy 165, 166
 - load 53
 - collaborative
 - distributed performance 260
 - virtual environment (CVE) 195, 196, 197, 205, 207, 208, 210
 - communication
 - channels 148
 - repertoires 148
 - communicative
 - act 145, 154, 230
 - behavior 145, 147, 149
 - component process approach 75, 83, 84, 85
 - computer games 54, 60, 67
 - Congruity Theory 224, 226, 227, 230
 - consciousness 109, 110
 - contingent goals 147, 148
 - control parameters 199, 200
 - conversational agent 156, 157, 158, 159
 - coping
 - abilities 75
 - potential 76
 - strategies 163
 - cultural factors 149
 - culture 145, 146, 148, 150, 152, 154, 156
- D**
- dancers 259, 261, 263, 266, 271, 273, 276
 - DBNs 125, 126, 136, 138
 - dementia 177, 179, 182, 185
 - depression 178, 190
 - dimension reduction 71
 - disabilities 258, 259, 268, 270, 271, 276, 277
 - discrete emotion 57, 59, 64
 - disembodied lady 111
 - disgust 204
 - display rules 47, 208
 - Distance Learning 196
 - dog 177, 178, 180, 182
 - dramatic interaction 171
 - Duchenne smile 41, 42, 43, 64
 - dynamic belief networks 125
 - dynamic presentation 64
- E**
- ecological fitness 5
 - elderly people 177, 179, 181, 182, 185, 186, 187, 188, 190, 191, 192
 - ELIZA 37
 - embodied
 - agents 21
 - character 123, 134
 - Embodied Conversational Agent (ECA) 48, 125, 139, 143, 145, 154, 156, 158, 173, 230, 232
 - embodiment 110
 - EMOTE 88, 89, 100
 - emotion
 - activation 123, 124, 126, 129, 131, 133, 134, 137, 138
 - recognition 58
 - emotional
 - animations 72
 - artifacts 105, 106
 - expression 71, 73, 74, 84, 86, 104, 105, 106
 - interfaces 54
 - problem solving 54
 - transforms 88, 92
 - emotions 144, 147, 154, 156, 158, 159, 160
 - emotive expressions 38
 - empathy 105, 116, 162, 165, 172, 196
 - envelope displays 239
 - ethnography 23, 24, 26, 35
 - evaluation results 162, 172
 - evolutionary biology 7
 - exaggeration 168
 - expression control mechanism 207
 - expressive
 - behaviour 238, 239, 242, 244
 - blend 44
 - displays 47, 48
 - movement 72, 79, 83
 - robots 38, 47, 48
 - eyebrows 55, 62, 63, 64
 - eye gaze 238, 247
- F**
- face-to-face conversations 235
 - Face Scale 187, 192
 - facial
 - configuration 39, 43
 - display 37, 40, 48, 50
 - expression(s) 37, 38, 39, 40, 43, 53, 66, 67, 123, 137, 139, 195, 196, 197, 198, 199, 201, 202, 204, 205, 207, 208
 - expression synthesis 53, 64
 - signals 55, 57
 - Facial Action Coding System (FACS) 56, 59, 60, 63, 64, 66, 67, 197, 199, 201, 207, 210
 - Facial Action Composing Environment 63, 68
 - Facial Expression Analysis Tool 60, 68
 - FantasyA 95, 97, 98, 100
 - fear 57, 58, 197, 201, 204, 205, 206
 - fictive personality 22, 27, 31
 - Flutterfugue 258
 - flying avatar 262
 - folk psychology 8
 - Forum Theatre 163
 - forward-kinematics 264
 - Foucault 24, 25, 26, 35, 36
 - Fourier Functional Model 88
 - friendship 12, 16
 - Furby 182
- G**
- gaming environment 35
 - gaze 106, 115, 116, 238, 239, 240, 242, 243, 247, 248, 249, 250, 252, 253
 - gender 47, 48, 173, 175
 - generative mechanism 125, 126
 - Geneva Appraisal Manipulation Environment 60, 68
 - gesticulation 149
 - gestural dialect 150
 - gestures 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158
 - GESTYLE 146, 150, 151, 152, 153, 156
 - goal 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 135, 136, 137, 138, 140
 - greetings 46
 - grounding problem 108
 - group selection 7, 8

H

H-Anim 151, 154, 199, 210
 happiness 124, 136
 hierarchical language 150
 household robots 280, 281,
 285, 289
 human movement 259

I

ideomotoric empathy 165
 intentionality 104, 106, 115
 Interactive Data Elicitation and
 Analysis tool 61
 interpersonal relations 14
 interpersonal relationships 226
 intra-personal meaning 249
 intrinsic properties 3
 inverse kinematics 83
 invisible friend 165

J

Jack toolkit 87
 joint-link model 90, 91
 Josephine 271, 274, 275

K

key-frame 91
 Kismet 37, 38, 47

L

Laban Movement Analysis 88
 lexical items 73
 linguistic control 242
 love 280, 283, 285, 287, 288,
 289, 291

M

Machiavellian intelligence 7, 18
 Magicster 125
 Marathon 124, 128, 130, 134, 135
 marionettes 261
 markup 245, 247
 languages 144, 156
 MARKUP 150, 151
 meeting organization 23, 31
 mental state(s) 4, 5, 59, 129, 131
 Microsoft Agent 28
 Mind-Tested toolkit 131
 mobile assistant 28
 modal emotions 59
 mood 29, 30

moral sentiments 3, 17
 motherese 283, 291
 motion characteristics 149,
 154, 155
 motivation system 38
 motor energy 148
 MPEG-4 145, 151, 153, 158
 multi-layered control 243
 multi-modal dialogs 146
 multi-party 228

N

narrative 21, 22, 23, 35
 naturalism 166, 168, 173
 natural selection 5, 7, 19
 negatively-valenced emotions
 127
 negotiation process 23
 neutral walk 89, 93, 95
 non-player characters (NPCs)
 235
 non-verbal
 behaviour 242, 244, 245
 channels 196
 communication 105, 237,
 243, 245
 signals 143, 146
 normative standards 75, 76,
 77, 84

O

Ortony, Clore and Collins
 (OCC) 123, 128
 confirmation 128, 130
 fortune-of-others 128
 happy-for 125, 128, 129, 130,
 131, 134, 136
 prospect-based 128, 130
 sorry-for 125, 128, 129, 130,
 131, 134
 well-being 128, 130
 Oz project 22, 23

P

PACEO 22, 27, 28, 29, 30, 31, 32,
 33, 34, 35, 237
 Paro 180, 181, 182, 183, 184, 185,
 186, 187, 188, 189, 190, 191
 passive stereo glasses 266
 Perlin Noise 88

Personal and Social Education
 162

Personalisation 245
 personality 124, 125, 129, 131,
 132, 133, 135, 137, 138, 139,
 141, 143, 144, 145, 147, 148,
 149, 150, 152, 157, 159, 160,
 174
 parameters 170

pets 177
 photo-realism 166
 physical therapy 269
 placebo 182, 185, 186, 188, 189,
 190, 191
 police 284, 285, 286, 287, 289
 positive affect 38, 44
 posture 71, 72, 78, 79, 80, 81,
 83, 168

power

relations 22, 26
 structure 23, 29, 34
 premeditated display 39

principle

of antithesis 6
 of direct action 6
 process modeling 108

Profile of Mood States 187, 188,
 189, 190, 191

proprioception 112

puppeteer 258, 264, 265, 266,
 267, 268, 269, 273

puppets 259, 260, 262, 263,
 266, 270

R

R781 280, 281, 282, 283, 284,
 285, 286, 287, 288, 289, 290,
 291

readout models 39

relational bullying 163, 173

repertoire 149, 155

robot 179, 180, 182, 183, 185, 192
 -assisted activities 181, 182
 performer 271

therapy 180

Robot Central 285, 287, 288,
 289, 290, 291

S

salient narratives 26

schadenfreude 44

- science fiction 279, 290
 - scowl 37
 - SEC 75, 77, 78, 79, 80, 81, 82, 83, 84, 85
 - selective autonomy 244
 - semi-autonomous avatar 235, 239, 240, 241, 244, 245, 250, 252
 - sense-reflect-act 169
 - SenToy 97, 98, 100
 - simulation 71, 72, 84, 85
 - situated
 - behavior 54
 - coding procedures 56
 - smile 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48
 - dimpler 45
 - duplay 43
 - play 43
 - smiling 38, 45, 46, 48, 50
 - SOAR 237
 - social
 - agents 103, 106, 113
 - animals 2, 4
 - awareness 22, 27
 - behaviour 123
 - bond 2, 3
 - construction of emotions 3
 - context 228, 229, 230
 - emotions 2, 3, 17
 - emotions thesis 2, 3, 17
 - interaction 22, 25, 28, 29, 30, 35, 103, 104, 106, 113, 114, 115, 117
 - message 40
 - obligation 46
 - organization 2
 - partners 105
 - robots 38
 - role 21, 22, 23, 33, 34
 - setting 148
 - standing 29
 - sociogram 223
 - spatial amplitude parameter 93
 - spherical linear interpolation 92, 93, 94
 - spontaneous smiles 39
 - stance 91, 92, 94, 95, 96, 99
 - STEVE 237
 - stimulus 74, 75, 76, 77, 78, 79, 80, 83, 85
 - evaluation 76
 - evaluation checks 59
 - strategic intraspecific
 - co-ordination 3, 12, 17, 18
 - communication 11, 12
 - relations 3, 12
 - style 143, 144, 145, 146, 147, 148, 149, 150, 154, 156, 157, 160
 - surprise 57, 58, 197, 201, 204, 205
 - synthesised speech 174
 - synthetic character 161
- T**
- Tamagotchis 165
 - telematic performance 259, 260
 - theory of mind 115, 161
 - threshold 149
 - tracking 239, 240, 242
 - Turing test 21
 - turn-taking 106, 228, 229
- U**
- Uncanny Valley 167, 198, 199
 - uncertainty 124, 126, 127, 136, 137
- universal facial expressions** 197, 200
- unscripted dramas** 169, 174
- user input** 241, 242, 246
- utility** 127, 136, 137
- V**
- valence 73, 74, 78
 - virtual
 - actors 173
 - butterfly 263, 266
 - character 236, 245
 - drama 162, 166
 - head 199, 200, 201, 202, 204, 205, 207, 208, 209
 - humans 143
 - puppetry 260
 - spectators 289
 - world 235, 236, 240, 241, 243
 - Virtual Interactive Puppetry 257, 259, 275, 276, 278
 - Virtual Reality Modeling Language (VRML) 199, 200
 - Virtual Teletubbies 237, 253
 - visual
 - displays 37
 - realism 88, 90, 99
 - volitional smiles 39
- W**
- working environments 27
- X**
- XML 170
- Z**
- zombies 167

Advances in Consciousness Research

A complete list of titles in this series can be found on the publishers' website, www.benjamins.com

- 75 **SKRBINA, David (ed.):** *Mind that Abides. Panpsychism in the new millennium.* xiv, 397 pp. + index. Expected February 2009
- 74 **CANAMERO, Lola and Ruth AYLETT (eds.):** *Animating Expressive Characters for Social Interaction.* 2008. xxiii, 296 pp.
- 73 **HARDCASTLE, Valerie Gray:** *Constructing the Self.* 2008. xi, 186 pp.
- 72 **JANZEN, Greg:** *The Reflexive Nature of Consciousness.* 2008. vii, 186 pp.
- 71 **KROIS, John Michael, Mats ROSENGREN, Angela STEIDELE and Dirk WESTERKAMP (eds.):** *Embodiment in Cognition and Culture.* 2007. xxii, 304 pp.
- 70 **RAKOVER, Sam S.: To Understand a Cat. Methodology and philosophy. 2007. xviii, 253 pp.**
- 69 **KUCZYNSKI, John-Michael:** *Conceptual Atomism and the Computational Theory of Mind. A defense of content-internalism and semantic externalism.* 2007. x, 524 pp.
- 68 **BRÄTEN, Stein (ed.):** *On Being Moved. From mirror neurons to empathy.* 2007. x, 333 pp.
- 67 **ALBERTAZZI, Liliana (ed.):** *Visual Thought. The depictive space of perception.* 2006. xii, 380 pp.
- 66 **VECCHI, Tomaso and Gabriella BOTTINI (eds.):** *Imagery and Spatial Cognition. Methods, models and cognitive assessment.* 2006. xiv, 436 pp.
- 65 **SHAUMYAN, Sebastian:** *Signs, Mind, and Reality. A theory of language as the folk model of the world.* 2006. xxvii, 315 pp.
- 64 **HURLBURT, Russell T. and Christopher L. HEAVEY:** *Exploring Inner Experience. The descriptive experience sampling method.* 2006. xii, 276 pp.
- 63 **BARTSCH, Renate:** *Memory and Understanding. Concept formation in Proust's *A la recherche du temps perdu*.* 2005. x, 160 pp.
- 62 **DE PREESTER, Helena and Veroniek KNOCKAERT (eds.):** *Body Image and Body Schema. Interdisciplinary perspectives on the body.* 2005. x, 346 pp.
- 61 **ELLIS, Ralph D.: Curious Emotions. Roots of consciousness and personality in motivated action. 2005. viii, 240 pp.**
- 60 **DIETRICH, Eric and Valerie Gray HARDCASTLE:** *Sisyphus's Boulder. Consciousness and the limits of the knowable.* 2005. xii, 136 pp.
- 59 **ZAHAVI, Dan, Thor GRÜNBAUM and Josef PARNAS (eds.):** *The Structure and Development of Self-Consciousness. Interdisciplinary perspectives.* 2004. xiv, 162 pp.
- 58 **GLOBUS, Gordon G., Karl H. PRIBRAM and Giuseppe VITIELLO (eds.):** *Brain and Being. At the boundary between science, philosophy, language and arts.* 2004. xii, 350 pp.
- 57 **WILDGEN, Wolfgang:** *The Evolution of Human Language. Scenarios, principles, and cultural dynamics.* 2004. xii, 240 pp.
- 56 **GENNARO, Rocco J. (ed.):** *Higher-Order Theories of Consciousness. An Anthology.* 2004. xii, 371 pp.
- 55 **PERUZZI, Alberto (ed.):** *Mind and Causality.* 2004. xiv, 235 pp.
- 54 **BEAUREGARD, Mario (ed.):** *Consciousness, Emotional Self-Regulation and the Brain.* 2004. xii, 294 pp.
- 53 **HATWELL, Yvette, Arlette STRERI and Edouard GENTAZ (eds.):** *Touching for Knowing. Cognitive psychology of haptic manual perception.* 2003. x, 322 pp.
- 52 **NORTHOFF, Georg:** *Philosophy of the Brain. The brain problem.* 2004. x, 433 pp.
- 51 **DROEGE, Paula:** *Caging the Beast. A theory of sensory consciousness.* 2003. x, 183 pp.
- 50 **GLOBUS, Gordon G.: Quantum Closures and Disclosures. Thinking-together postphenomenology and quantum brain dynamics. 2003. xxii, 200 pp.**
- 49 **OSAKA, Naoyuki (ed.):** *Neural Basis of Consciousness.* 2003. viii, 227 pp.
- 48 **JIMÉNEZ, Luis (ed.):** *Attention and Implicit Learning.* 2003. x, 385 pp.
- 47 **COOK, Norman D.: Tone of Voice and Mind. The connections between intonation, emotion, cognition and consciousness. 2002. x, 293 pp.**
- 46 **MATEAS, Michael and Phoebe SENGERS (eds.):** *Narrative Intelligence.* 2003. viii, 342 pp.
- 45 **DOKIC, Jérôme and Joëlle PROUST (eds.):** *Simulation and Knowledge of Action.* 2002. xxii, 271 pp.
- 44 **MOORE, Simon C. and Mike OAKSFORD (eds.):** *Emotional Cognition. From brain to behaviour.* 2002. vi, 350 pp.
- 43 **DEPRAZ, Nathalie, Francisco J. VARELA and Pierre VERMERSCH:** *On Becoming Aware. A pragmatics of experiencing.* 2003. viii, 283 pp.

- 42 STAMENOV, Maxim I. and Vittorio GALLESE (eds.): Mirror Neurons and the Evolution of Brain and Language. 2002. viii, 392 pp.
- 41 ALBERTAZZI, Liliana (ed.): Unfolding Perceptual Continua. 2002. vi, 296 pp.
- 40 MANDLER, George: Consciousness Recovered. Psychological functions and origins of conscious thought. 2002. xii, 142 pp.
- 39 BARTSCH, Renate: Consciousness Emerging. The dynamics of perception, imagination, action, memory, thought, and language. 2002. x, 258 pp.
- 38 SALZARULO, Piero and Gianluca FICCA (eds.): Awakening and Sleep–Wake Cycle Across Development. 2002. vi, 283 pp.
- 37 PYLKKÄNEN, Paavo and Tere VADÉN (eds.): Dimensions of Conscious Experience. 2001. xiv, 209 pp.
- 36 PERRY, Elaine, Heather ASHTON and Allan H. YOUNG (eds.): Neurochemistry of Consciousness. Neurotransmitters in mind. With a foreword by Susan Greenfield. 2002. xii, 344 pp.
- 35 Mc KEVITT, Paul, Seán Ó NUALLÁIN and Conn MULVIHILL (eds.): Language, Vision and Music. Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, 1999. 2002. xii, 433 pp.
- 34 FETZER, James H. (ed.): Consciousness Evolving. 2002. xx, 253 pp.
- 33 YASUE, Kunio, Mari JIBU and Tarcisio DELLA SENTA (eds.): No Matter, Never Mind. Proceedings of Toward a Science of Consciousness: Fundamental approaches, Tokyo 1999. 2002. xvi, 391 pp.
- 32 VITIELLO, Giuseppe: My Double Unveiled. The dissipative quantum model of brain. 2001. xvi, 163 pp.
- 31 RAKOVER, Sam S. and Baruch CAHLON: Face Recognition. Cognitive and computational processes. 2001. x, 306 pp.
- 30 BROOK, Andrew and Richard C. DEVIDI (eds.): Self-Reference and Self-Awareness. 2001. viii, 277 pp.
- 29 VAN LOOCKE, Philip (ed.): The Physical Nature of Consciousness. 2001. viii, 321 pp.
- 28 ZACHAR, Peter: Psychological Concepts and Biological Psychiatry. A philosophical analysis. 2000. xx, 342 pp.
- 27 GILLET, Grant R. and John McMILLAN: Consciousness and Intentionality. 2001. x, 265 pp.
- 26 Ó NUALLÁIN, Seán (ed.): Spatial Cognition. Foundations and applications. 2000. xvi, 366 pp.
- 25 BACHMANN, Talis: Microgenetic Approach to the Conscious Mind. 2000. xiv, 300 pp.
- 24 ROVEE-COLLIER, Carolyn, Harlene HAYNE and Michael COLOMBO: The Development of Implicit and Explicit Memory. 2000. x, 324 pp.
- 23 ZAHAVI, Dan (ed.): Exploring the Self. Philosophical and psychopathological perspectives on self-experience. 2000. viii, 301 pp.
- 22 ROSSETTI, Yves and Antti REVONSUO (eds.): Beyond Dissociation. Interaction between dissociated implicit and explicit processing. 2000. x, 372 pp.
- 21 HUTTO, Daniel D.: Beyond Physicalism. 2000. xvi, 306 pp.
- 20 KUNZENDORF, Robert G. and Benjamin WALLACE (eds.): Individual Differences in Conscious Experience. 2000. xii, 412 pp.
- 19 DAUTENHAHN, Kerstin (ed.): Human Cognition and Social Agent Technology. 2000. xxiv, 448 pp.
- 18 PALMER, Gary B. and Debra J. OCCHI (eds.): Languages of Sentiment. Cultural constructions of emotional substrates. 1999. vi, 272 pp.
- 17 HUTTO, Daniel D.: The Presence of Mind. 1999. xiv, 252 pp.
- 16 ELLIS, Ralph D. and Natika NEWTON (eds.): The Caldron of Consciousness. Motivation, affect and self-organization — An anthology. 2000. xxii, 276 pp.
- 15 CHALLIS, Bradford H. and Boris M. VELICHKOVSKY (eds.): Stratification in Cognition and Consciousness. 1999. viii, 293 pp.
- 14 SHEETS-JOHNSTONE, Maxine: The Primacy of Movement. 1999. xxxiv, 583 pp.
- 13 VELMANS, Max (ed.): Investigating Phenomenal Consciousness. New methodologies and maps. 2000. xii, 381 pp.
- 12 STAMENOV, Maxim I. (ed.): Language Structure, Discourse and the Access to Consciousness. 1997. xii, 364 pp.
- 11 PYLKKÖ, Pauli: The Aconceptual Mind. Heideggerian themes in holistic naturalism. 1998. xxvi, 297 pp.
- 10 NEWTON, Natika: Foundations of Understanding. 1996. x, 211 pp.
- 9 Ó NUALLÁIN, Seán, Paul Mc KEVITT and Eoghan Mac AOGÁIN (eds.): Two Sciences of Mind. Readings in cognitive science and consciousness. 1997. xii, 490 pp.

- 8 **GROSSENBACHER, Peter G. (ed.):** Finding Consciousness in the Brain. A neurocognitive approach. 2001. xvi, 326 pp.
- 7 **MAC CORMAC, Earl and Maxim I. STAMENOV (eds.):** Fractals of Brain, Fractals of Mind. In search of a symmetry bond. 1996. x, 359 pp.
- 6 **GENNARO, Rocco J.:** Consciousness and Self-Consciousness. A defense of the higher-order thought theory of consciousness. 1996. x, 220 pp.
- 5 **STUBENBERG, Leopold:** Consciousness and Qualia. 1998. x, 368 pp.
- 4 **HARDCASTLE, Valerie Gray:** Locating Consciousness. 1995. xviii, 266 pp.
- 3 **JIBU, Mari and Kunio YASUE:** Quantum Brain Dynamics and Consciousness. An introduction. 1995. xvi, 244 pp.
- 2 **ELLIS, Ralph D.:** Questioning Consciousness. The interplay of imagery, cognition, and emotion in the human brain. 1995. viii, 262 pp.
- 1 **GLOBUS, Gordon G.:** The Postmodern Brain. 1995. xii, 188 pp.

